INEQUALITY OF YEARS OF SCHOOLING, STUNTING, AND INTRA-HOUSEHOLD BMI IN CHINA

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Xiaoxiao Ma

August 2019

ABSTRACT

Despite the rapid economic growth in China, income inequality has widened, especially between the urban, eastern coastal provinces and rural inland areas due to unbalanced economic development and different access to resources. Inequality in education and health are correlated with income inequality, and in turn affects people's human capital and ability to work, further reinforcing the income inequality. In Chapter I, I explore to what extent is inequality of years of schooling in China from half a century ago to nowadays affected by circumstance factors rather than personal effort. I decompose inequality into between group inequality (inequality of opportunity which is determined by circumstances) and within group inequality. I apply China Family Panel Studies (CFPS) 2016 follow up survey data to compare the results with the previous study and discuss relative policy implications. I find inequality for the opportunity constitutes 29% of the educational outcome inequality on population born from 1940 to 1994 and hukou is the most influential circumstance factor. In Chapter II, I investigate the environmental factors that cause child stunting inequality in China. I apply China Health and Nutrition Survey (CHNS) data from 1989 to 2011 to measure stunting inequality by concentration index and decompose changes in inequalities. I find the notable income growth in China reduces average stunting level. However, increased income exaggerates stunting inequality by increasing income inequality's contribution to stunting inequality. In Chapter III, I explore to what extent is inequality well-being as represented by body mass index (BMI) attributable to intra-household inequality, whether the health resource deprivation exists among family members, and whether a

Kuznets curve relationship between intra-household inequality and average well-being exists. From available data from CHNS, I find intra-household inequality accounts for more than a half of outcome inequality, and no clear pattern of deprivation among family members nor Kuznets curve relationship has been found.

BIOGRAPHICAL SKETCH

Xiaoxiao Ma completed bachelor's degree of Science majoring in Agribusiness Management from Pennsylvania State University in May 2016.

ACKNOWLEDGMENTS

I would like to express my very great appreciation to Professor Kanbur and Professor Nicholson, my research supervisors, for their patient guidance, enthusiastic encouragement and useful critiques of this research work. Their willingness to give their time so generously has been very much appreciated.

I would also like to thank the staff of the Cornell Statistical Consulting Unit for their valuable and constructive suggestions during the development of this research work.

This research has not been funded by outside grants.

TABLE OF CONTENTS

CHAPTER I

INEQUALITY OF OPPORTUNITY FOR YEARS OF EDUCATION IN CHINA FROM 1940 TO 1994

Introduction

Beginning with the end of the 20th century, China is experiencing a social transformation with unprecedented influence, speed, and scope. China's economy has expanded rapidly and stably since the economic reform and opening was initiated 1978. The per-capita GDP attaied an annual growth of 6.7%[1] from 1978 to 2008. Educational attainment has also markedly increased in China, especially at the postsecondary level. The college enrollment rate boomed in the late 1990s, and the rapid increase of educational attainment is both the result of the economic development, and fuels continuing growth.

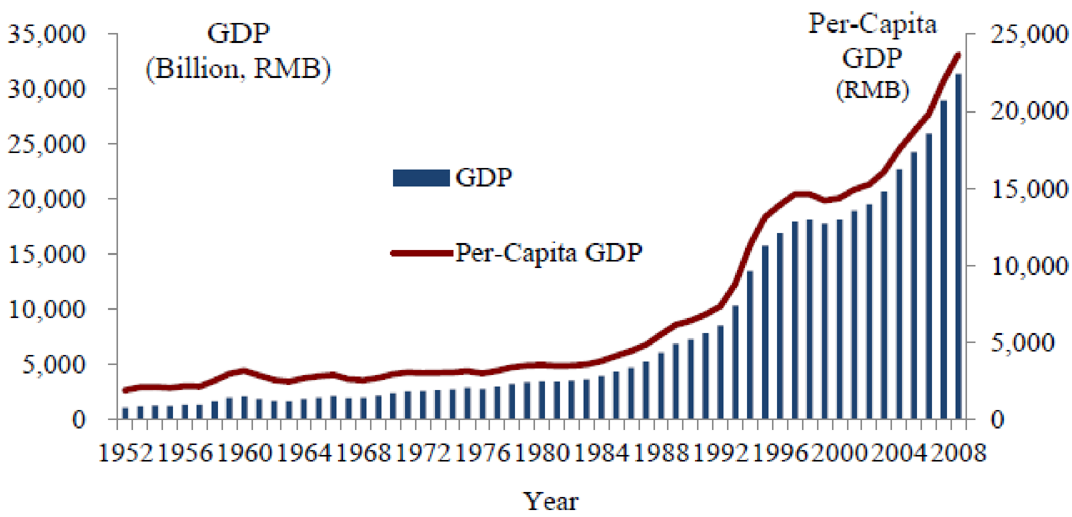


Figure I.1(a). Trends in GDP and per-capita GDP, 1952-2008 (in 2008 RMB)
Note: Adjustment has been made for the data of 2005-2008, on the basis of the 2d Economic Census.
Source: Stat Bureau of Statistics. 2010. China Statistical Abstracts 1949-2008. China Statistics Press.

[1] GDP growth rate without inflation, China Statistical Yearbook, National Bureau of Statistics of China

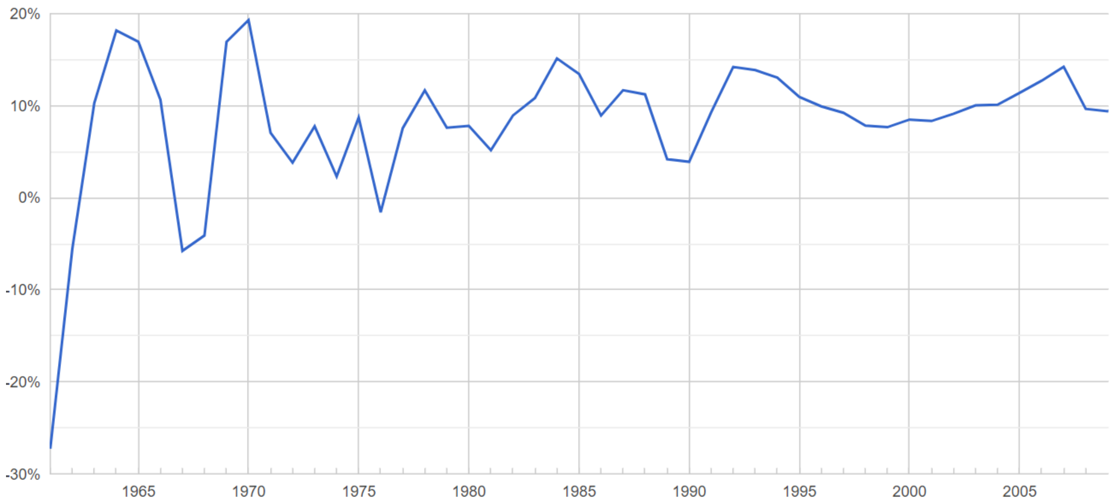


Figure I.1(b). Percentage change of China real GDP 1961-2008.
Source: World Bank

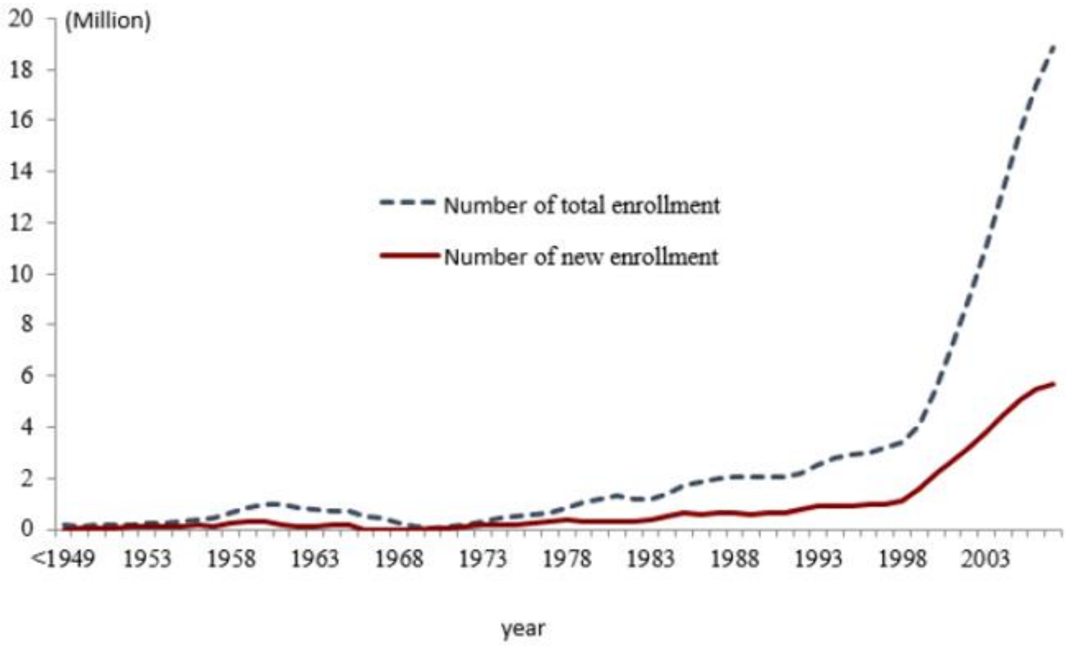


Figure I.2. Higher education expansion in China, 1949-2007.
Source: China Education Yearbook Editorial Board (1984, 1986-1988, 1989-2008)

In China, education is considered the fairest way for children rising from different socioeconomic backgrounds to compete in the job market and pursue success. Through the ages, education is one of the most concerning topics in politics that it relates the fate of individuals to the long-term development of the whole country (Benos and Zotou, 2004; Kunofiwa, 2017; Delalibera and Ferreira, 2019). Therefore, Chinese government takes educational affairs seriously and funds expended on education constitute a substantial portion of the budget. Despite the heavy investment, social mobility is slowing down, and individual effort seems to have less impact on educational attainment. The interview of the top scorer in 2017 college entrance examination (Liberal arts, Beijing area) drew widespread criticism when he considered himself benefiting from the privileged family socioeconomic status and enjoying the “shortcut” of living in Beijing and accessing to the best educational resources in China.

Therefore, despite the overall outcome improvement, it’s also essential to look into inequality during the process. Equity theory (Adams, 1963) points out that individuals are motivated by fairness, and if they identify inequities of themselves, they will seek to adjust their input to reach their perceived equity, and hence influence the political stability and economic development. In Chinese history, the Cultural Revolution was an extremally radical approach to pursue equal redistribution, and it resulted in severe economic and educational recession. As many political philosophers (Rawls, 1958) considered, inequality is acceptable only when it results from different levels of effort, a personal responsibility. A concept in developing economics “equality of opportunity” (EOp) is developed by economists including Dworkin (1981), Arneson (1989), Cohen (1989), Roemer (1993, 1998) and Checchi and Peragine (2010). The central idea of this concept is dividing the causes of inequality into fair causes and unfair causes, in this case, effort and circumstance. Populations with the same circumstances are grouped into finite types composed by circumstance factors such as gender, ethnicity, family income, parental education, etc. Outcome is presented by a utility function determined by effort and a set of social policies, which is considered the average achievement obtained among people in a typical type. The inequality in outcome can be decomposed into between group inequality and within group inequality. Within group inequality is determined by effort and policy since each group has identical circumstance factors and the between group inequality represents the inequality of opportunity.

Another reason to look into inequality is its relationship between the overall outcome mean. Kuznets (1950) proposed a hypothesis that as economy develops, the inequality first increase and then decreases, forming an inverted-U shape curve. So even the mean outcome is improving, it is possible that the population is more unequal, and a growing proportion of the population fall below the minimum threshold (a “poverty line”). The third reason to investigate inequality is that inequality in one indicator may be causal for other forms of inequality (for example, in Chapter III we discuss stunting inequality will further affect educational attainment and income).

The literature on measuring inequality of opportunity has many theoretical and empirical applications on various topics including income, education, health etc. (De Barros et al., 2009; Van de Gaer, 1993; Ferreira and Gignoux 2011) Applications in China include Golley and Kong (2018) in education, Zhang and Kanbur (2005) on spatial inequality in education and healthcare, and Zhang and Eriksson (2010) in income. Golley and Kong’s work find inequality of opportunity contributes 33.4%~39.3% to total educational inequality. Zhang and Kanbur’s work point out increasingly polarized urban-rural illiteracy rate from 17.8 in 1981 and 25.7 in 2000.

As for health outcomes, the gap between rural-urban infant mortality rate rose from 1.5 in 1981 to 2.1 in 2000. Zhang and Eriksson discover income inequality of opportunity rises continuously from 46% to 63% in 1989-2006.

Previous researches to explore the educational inequalities in China include Liu (2006), Wu and Huang (2015), and Zhou et al. (1998). Liu (2006) uses a Cox proportional hazard model to assess the possibility of people from different class backgrounds got enrolled into higher education. Wu and Huang (2015) show Blinder Oaxaca decompositions of gender disparities in educational outcomes. Zhou, Moen and Tuma (1998) focus on urban China and uses logistic regression model to calculate the probability of attaining each level of education (junior high, senior high and college) across class backgrounds, gender, region and family size. However, none of the information currently available indicates how educational inequality of children with different circumstances has changed over time and how changing circumstance factors have impacted on the distribution of educational outcomes. Recent research by Golley and Kong (2018) analyzes data from the China Family Panel Studies (CFPS) survey in 2010 and 2012 to investigate the influence of circumstance factors on education outcome inequality represented by years of schooling. They also explore how the proportion of inequality caused by circumstance factors change by time. The purpose of this article is to explore circumstances factors and examine to what extent does inequality of years of schooling in Chins from half a century ago to nowadays affect by circumstance factors rather than personal effort. I build my work on their study and apply CFPS 2016 follow up survey data to compare the results and discuss relative policy implications.

Literature

a. Inequality of opportunity

Roemer's study (1993, 1998) was considered one of the foundations of measuring opportunity of equality. A generalized entropy model was applied to calculate "accountable effort". Populations with identical circumstances are grouped into finite types composed by circumstance factors such as gender, ethnicity, parental political status, parental education, etc. Ideally, in each type, the advantage outcome y should be determined only by personal effort. The decomposition of total inequality can be written as:

$$E_\alpha(y) = E_\alpha(\{\mu_k^i\}) + \sum_{k=1}^{K} \frac{n^k}{N} \left(\frac{\mu^k}{\mu}\right)^\alpha E_\alpha(y^k) \tag{1}$$

where n^k and y^k denote, respectively, the population and the advantage distribution in type k, and α is the generalized entropy parameter. The first term in the right-hand side of this equation—the between-group component—is inequality in the smoothed distribution. The second term is the within-group component.

One remarkable achievement from Roemer's work was to recognize some part of effort is determined by circumstance factors, like policy taken in a particular group. He defines the distribution of effort (e) within each type $G_\rho^k(e)$ conditional on certain policy ρ, and denotes the advantage level y enjoyed by a person in quantile π $=G_\rho^k(e)$ in the effort distribution of type k, given policy ρ, as $y^k(\pi,\rho)$. Roemer proposed that individual located at the same centile among within group distributions should receive same outcome: $y^k(\pi,\rho) = y^l(\pi,\rho)$. Assuming a monotone increasing utility function for personal effort, “opportunity-equalizing policy” can be solved from maximizing the average outcome of most disadvantaged group:

$$\rho^* = argmax \int_0^1 miny^k\,(\pi, \rho) d\pi$$

The “strong criterion” definition of equal opportunity requires identical distribution across types, while the “weak criterion” definition only expects the mean outcome level to be identical across types (Van de Gaer, 1993). Lefranc et al. (2008) applies stochastic dominance rankings conditional on types under strong criterion and claim the strong criterion only holds when no type stochastically dominants another type. The weak criterion calculates inequality of opportunity as the between group inequality when suppressing the within group inequality. The sample size required in weak criterion is less than strong criterion and allows to consider more circumstances empirically. This study was the foundation of many other related studies conducted in various subjects.

Other than decomposing total inequality into within and between group inequalities (the later term is referred to inequality of opportunity), Bourguignon, Ferreira, and Menéndez (2007a) consider outcome variable a function of circumstances and effort and estimate the importance of opportunity-forming circumstances in accounting for earnings inequality. The circumstance component is further decomposed into a direct effect and an indirect effect that operates through the influence of circumstances on the choice of efforts. They decompose overall earnings inequality in Brazil into race, parents’ education, father’s job and location. They also estimate the lower bound of the measurement considering the unobserved circumstances. The observed circumstances are found to account for between 10%~37% of the overall inequality.

A substantial amount of literature has adopted the concept of inequality of opportunity and apply it in a wide range of economic outcomes: income, expenditure, education, health, etc. For example, Ferreira and Gignoux (2011) construct a scalar measure of inequality of opportunity and apply to six Latin American countries. The study applies comparison between within-group inequality and between-group inequality, where between-group inequality was considered unfair inequality due to predetermined circumstances. The findings reported that inequality of opportunity constitutes ⅕ ~ ⅓ of total income inequality and ¼ ~ ½ of total consumption inequality. The study implied that the differences between nonparametric and parametric methods are statistically irrelevant.

Far less has been done in measuring inequality of opportunity in education area. One application in educational inequality of opportunity is from De Barros et al. (2009). They use PISA (Programme for International Student Assessment) standardized score in Latin American and conclude inequality of opportunity contributes 14~28% of the total outcome inequality, explained by circumstance variables including parents' education, father's occupation, gender and location. Also applying PISA score, Ferreira and Gignoux (2000, 2003) find 27% ~ 33% of educational outcome inequality in Turkey and 35% of all educational disparity world widely is attributable to inequality of education. Golleg and Kong's work (2018) is the first attempt to apply years of schooling as outcome of interest.

b. Political implications

Why does inequality of opportunity matter? Lv (2013) finds that public perception of social fairness, especially equal access to education, positively affects the acceptance level of income inequality and the attitude towards redistribution. Education is one type of public good that affected largely by policies, and it is excludable to some extent. The disparity in education could occur in geographic level and individua level. Specially, eastern and coastal areas of China enjoy superior educational resources than central and western part of China due to historically imbalanced economic development. In each Province, their different entitlement rules limits admission of students from different locations, which usually favors local students than students out of province. Also, Hukou policy differentiates students with urban and rural hukou, which students with rural hukou usually are not able to enjoy same resources as urban students do even if they live in urban areas. Lv considers equal education opportunity an effective filter when assessing another person's success, by placing everyone at the same starting line in a way. Because people typically don't have details about how others achieve their success, the perception of equal opportunity, in this case, equal education opportunity, largely determines their assessment of whether others deserve their success or not, in turn determines their approval of redistribution policies among their society. The dissatisfaction towards redistribution policies and resentment towards income inequality would potentially impair social stability and hinder social development. Therefore, policies equalizing opportunities could help assuage the trouble of rising income inequality. And it should be the inequality of opportunity, not the inequality of outcomes that influence the design of such policies.

The political agenda is highly correlated with China's educational outcomes. Knight, Sicular and Yue (2011) use CHIP 2007 survey data to examine the intergenerational dimension of educational inequality. The survey sample covers individuals born from 1930s to 1980s, a period when China experienced huge economic and social change, including educational policies. When the People's Republic of China was established in 1945, one goal was to reduce illiteracy, so universal education was popularized rapidly. In the First Five Year Plan (1953~1957), resources were devoted to secondary education to train skilled workers, as a part of Soviet communism strategy to develop heavy industry. From this phase, educational funding was inclined to urban areas, while rural schools were typically funded by local farm cooperatives. The Great Leap Forward in 1958 shifted education agenda with political agenda to "left", a more aggressive communism model. Enrollment into primary level of education grew rapidly while curriculums showed more political tendency. By around 1960, some state funded key schools were formed, providing high quality education to those intelligent students with good political background. During the decade of Cultural Revolution (1966-1976) policies shifted to extreme left and universities were closed. Mass education was promoted with collective farming while elitism was disdained. Gaokao (national entry test to universities) was not resumed until 1977. Along with reform and opening policy starting from 1978, educational goals shifted from mass education to high quality education. Since then, China's economy has grown dramatically, but with increasing inequality. Rich areas could better subsidize local schools whereas poor areas fell behind. Knight, Sicular and Yue find that from 1980s, educational inequality declines with rising overall education level, measured by Gini coefficient. Parental education

$\bar{K}$.

inequality only accounts for less than 20% of children's educational inequality, and rural areas have lower contribution than urban. They also believe that class based social discrimination plays an important role in the drop out ratio, that is, children from poor family are more likely to drop out of school. Besides, studies in other countries find positive correlation between parental education and children's education even if education is heavily subsidized

Methods

We follow Roemer' approach and apply a generalized entropy model (equation 1) to calculate inequality. We also follow Golley and Kong's article and treat years of schooling as economic outcome of interest. Because we use years of schooling as the key variable, we can only use GE(2) to deal with those individuals with 0 year of schooling. GE(0) and GE(1) that involve taking log value are more often used in income inequality decompositions (Golley and Kong, 2011).

The approach is developed as follows. Consider a finite population of individuals, $\mathrm{i} \in \{1,2,\dots,\mathrm{N}\}$, and each individual achieves educational attainment y_i, which is a function of vector on circumstances C_i and effort e_i. C_i has J elements and each element of C_i^J takes a finite value x_J. Following Roemer (1998)'s approach and dividing the population with identical circumstances into *K* groups (types), and the partition $\pi \in \{T_1, \dots, T_K\}$, which distribution of y_i^k in each type *k* has identical circumstances: $C_i = C_j, \forall i, j / i \in T_k, j \in T_k$. The maximum number of possible types $\bar{K} = \prod_{J=1}^{J} x_j$.

We want to derive the inequality caused by difference in circumstances. The outcome inequality measured in generalized entropy class can be decomposed into between group inequality and within group inequality. Between group inequality can be also written as: $E_0(\{\mu_i^k\}) = \frac{1}{N}\sum_{i=1}^{N} log\frac{\mu}{\mu_i^k}$. Considering each individual inside the "Roemerian" type has identical circumstances, the inequality caused by different level of effort is within group inequality while inequality of opportunity is between group inequality.

The *direct* approach to measure the *absolute* scale of inequality of opportunity we suppress the within group inequality by assigning each individual the group mean and then measure the inequality of the smoothed distribution. $IOA_D = \mathrm{I}\{(\mu_i^k)\}$ where μ_i^k is the smoothed distribution that replacing each outcome y_i^k with group-specific mean μ_k. The *relative* measurement is the proportion of between group inequality over outcome inequality: $IOR_D = \mathrm{I}\{(\mu_i^k)\}/\mathrm{I}\{y_i\}$.

The *indirect* approach suppresses the between group inequality by replacing individual outcome y_i^k with $y_i^k\frac{\mu}{\mu_k}$, where μ stands for the grand mean and generates a standardized distribution v_i^k, which only leaves the within group inequality $\mathrm{I}\{v_i^k\}$. Thus, between group inequality, also the *indirect* measurement of *absolute* inequality of opportunity is $IOA_I=I(\{y_i\}-I(\{v_i^k\}$. similarly, the *indirect relative* term of inequality of opportunity is $IOR_I=1-I(\{v_i^k\}/I(\{y_i\})$.

Foster and Shneyerov (2000) find only GE(0) satisfies path-independent decomposability axiom[2] hence can achieve $IOA_D=IOA_I$, and $IOR_D=IOR_I$.

[2] Just as a smoothed distribution eliminates all within-group inequality by construction, a standardized distribution eliminates all between-group inequality, by appropriately rescaling all subgroup means. One might wish to impose the requirement that $\mathrm{I}(\{\mu_i^k\}) = \mathrm{I}(y) - \mathrm{I}(\{v_i^k\})$. This requirement is the axiom of path independent decomposability.

Applying OLS regression, the parametric analogue of inequality of opportunity measurement considers the advantage variable a function of circumstances, effort, and random unobserved circumstances y=f(C,e,u). While circumstances are defined as exogenous, effort can be influenced by circumstances and other factors. Thus, the expression can be written as y=(C,e(C,v),u). The reduced-form regression can be written as:

$$y = \beta C + \varepsilon \tag{2}$$

where y is year of schooling, C is a vector of circumstance variables and ε is the error term. We construct the prediction distribution with coefficients β and actual C: $y_i = C_i\beta$, with y replaced by its prediction. Since individual in the same group has identical circumstance vectors, the within group inequality is eliminated. Hence, we get the *direct* measurement of inequality of opponent: $IOA_D = I(\{y\})$ and $IOR_D = I(\{y\})/I(\{y_i^k\})$. We can also construct the standardized distribution with coefficients β, mean of circumstances $\widehat{C}$, and counting within-group variation ε_i: $\widehat{y_i} = \hat{C}\beta + \varepsilon_i$. Replacing C with $\widehat{C}$ eliminates between group inequality and only leaves within group inequality. Hence, we have $IOA_I = I(\{y_i^k\} - I(\{\hat{y}\})$ and $IOR_I = 1 - I(\{\hat{y}\}/I(\{y_i^k\})$ for indirect approach. We can also calculate partial contribution to total outcome inequality by suppressing circumstance with its mean value once at a time, and then calculating the remaining inequality in absolute and relative terms. Larger the partial inequality remains; smaller the circumstance's contribution is. Golley and Kong's article call such a distribution "counterfactual distribution": $y_i^J = C_i^{j \neq J}\beta_i + C^{-j=J}\beta_i$, and $IOA_P = I(\{y^J\}), IOR_P = I(\{y^J\})/I\big(\{y_i^k\}\big)$.

To test the adequacy of OLS model, we need to test for the normality, multicollinearity, model specification, and homoscedasticity. Normality of residuals is only required for valid hypothesis testing and is not required in order to obtain unbiased estimates of the regression coefficients. The term collinearity implies that two variables are near perfect linear combinations of one another. As the degree of multicollinearity increases, the regression model estimates of the coefficients become unstable and the standard errors for the coefficients can get wildly inflated. A model specification error can occur when one or more relevant variables are omitted from the model or one or more irrelevant variables are included in the model. If relevant variables are omitted from the model, the common variance they share with included variables may be wrongly attributed to those variables, and the error term is inflated. On the other hand, if irrelevant variables are included in the model, the common variance they share with included variables may be wrongly attributed to them. Model specification errors can substantially affect the estimate of regression coefficients. OLS assumes that the variance of the error term is constant (homoskedasticity). If the error terms do not have constant variance, they are said to be heteroskedastic. In our case we expect heteroskedasticity to present because there are subpopulation differences (Williams, 2015[3]), which violate the assumption of OLS. However, heteroskedasticity does not result in biased parameter estimates. Violation of the OLS assumptions will result in overestimated confidence intervals and underestimated p-value, and thus the statistical importance level of the independent

[3] Richard Williams, University of Notre Dame, https://www3.nd.edu/~rwilliam/ Last revised January 30, 2015, page 2

variable may also be overestimated.

Ferreira and Gignoux (2011) consider IOA and IOR the lower-bound of the inequality measurement because there probably exist some unobservable circumstances that influence the outcome inequality. The lower-bound result holds regardless the relationship between circumstances and effort. In their example, hours of study can be treated as unobserved effort which has no correlation with circumstance variables; or the effort corelated to circumstances (rural students study more in order to catch up/study less to help with housework), which will reflect in the biased weight (β) of observed circumstance: hukou. Therefore, we need to be cautious when interpreting the coefficients.

Data

a. CFPS

Lead by the Peking University research team, the China Family Panel Studies (CFPS) is an integrated national social survey that collects data on individual, family and community level. The project is implemented by the Institute of Social Science Survey (ISSS) at Peking University, assisted by University of Michigan and many other domestic government agencies , and funded by National Natural Science Foundation of China and Peking University. CFPS focus on economic and non-economic well-being of Chinese people. CFPS collects multi-level data to shape an individual's life by building relationships inside the society. In the community level, political environment, infrastructure, access to resources, transportation, medical care, financial revenue and expenditure are collected. At the household level data are collected on family structure, living conditions, social interaction, income and expenditure, composition of assets. Topics on education, occupation, marriage, income, physical and mental health are collected in the individual level. For children under 10 and family member who is not present, a proxy questionnaire is applied. The panel study tracks changes over time and allows researchers to follow the (dis)continuity of social phenomena.

CFPS collects data in 25 provinces[4] (municipalities/autonomous regions) that cover almost 95% of the Chinese population, hence is considered nationally representative. CFPS implemented Probability-Proportional-to-Size Sampling (PPS)[5] with implicit stratification.

In the 2010 baseline survey, CFPS interviewed 14,960 households, 33,600 adults and 8,990 adolescents with an approximate response rate of 82%. The 2010 baseline survey uses face-to-face interviews and collects topics on economic activities, educational attainment, family relationship and dynamics, mental and physical health, etc. The full sample follow-up surveys (2012, 2014, 2016) conducted every 2 years after 2010 use both face-to-face and telephone interview to trace, adjust and update information on gene[6] members interviewed in 2010 baseline survey. The follow-up survey also traces

---

[4] Half of the sample was generated by oversampling with five independent sampling frames (called —large provinces‖) of Shanghai, Liaoning, Henan, Gansu, and Guangdong. The other half households were from an independent sampling frame composed of 20 provinces (called —small provinces‖). Xie&Lu, 2015

[5] CFPS user guide: "Administrative units and socioeconomic status (SES) were used as the main stratification variables. Within the administrative unit, local GDP per capita was used as the ordering index for SES. If the GDP per capita in the administrative unit is not available, the proportion of nonagricultural population or population density is used."

[6] CFPS user guide: "All family members who were identified at baseline to have blood/marital/adoptive ties with the household were identified as CFPS gene members. In the follow-up surveys, newly born or adopted children of gene members were also considered CFPS gene members." For non-gene members the individual survey will be terminated.

proxy questionnaire from the last round of survey. I use 2010 baseline survey and 2012-2016 follow up survey data in this study. Many survey questions I use in this study, such as parents' education, are not included in the baseline survey but in the follow-up surveys.

b. Educational outcome indicator

I use years of formal schooling as the indicator of educational outcome, but due to unequal educational resources in different geographic locations and levels of schools, same years of education received do not necessarily mean equal education received. For example, colleges in China is divided into first tier, second tier and third tier (like community colleges in the U.S.), and inside the first-tier colleges there are "Project 985", "Project 211" and normal colleges. Students from the "Project 985" colleges are expected to receive the top-level educational resources and most popular in the job market with the highest potential income. The divergence between educational quality of the rural-poor and the urban elites further deepens the inequality in future competition. Oppedisano and Turati (2015) use standardized PISA score to measure the inequality of educational outcome in European countries. The standardized score can reflect students' study ability, and it also reflects the level of personal effort with random error, hence it seems to be a better indicator. However, there are no data available at national level that reflect students' learning ability and collecting such data requires significant amount of effort, considering China uses different teaching materials and examination systems in different regions. Likewise, the national college entrance exam score is infeasible because different enrollment policies are applied to students from different regions and it excludes students who only receive senior high school and below level of education. In addition, a small but important subgroup is missed in the sample: the students who study abroad. According to the 2017 Studying Abroad Development and Trend Report[7], there are about 0.6 million students who studied abroad in 2016, accounting for 3% of the college enrollment number in the same time period. Though the subgroup is small compared to the total population, this subgroup stands for the students with either the highest study ability (those who survive the intense competition and receive national fund to study abroad), and/or the most advantageous family socioeconomic status (considering the cost of studying abroad). Not including this subgroup, we truncate the right tail of the distribution. Because we treat individuals receiving same years of education as a group, we could underestimate the within group inequality, so the "lower bound" estimation in this article has a certain reference value.

c. Circumstances

Many studies show that the most influential factor that causes and reinforces inequality in China is the unique hukou system. (Golley and Kong, 2012, 2013; Knight, Sicular, and Yue, 2012; Li, et al., 2015; Luo et al., 2011; Qiao, 2008; Wang et al., 2009; Wu, 2009 and Yi et al., 2012). Hukou is a household registration system that prevents free movement of population and limits individual's activity inside a certain geographic area. Hence, it influences individual's access to pension fund, educational and health resources, especially between rural and urban division. In addition, intergenerational

[7] Retrieved from http://www.eol.cn/html/lx/report2017/yi.shtml (in Chinese)

persistence in educational attainment can be realized by combination of better parenting skills, transmission of intelligence, more investment on education and better economic status positively correlated with parents' higher educational attainment (Black and Devereux, 2011). During specific periods (during and after Mao's era, 1945~1976), family class status (Chengfen, including proletariat, worker, landlord, Bourgeois, etc.) plays an important role in accessing educational resources. During the cultural revolution, educational policies discriminated against the families with high socioeconomic status. The class labels are abandoned in 1979, but we can still measure the effect of class status by parents' Communist party membership. Much literature shows other circumstances not under an individual's control, including sex (Zhang et al., 2012; Zhang and Chen, 2014), ethnic groups (Li et al., 2015; Yang and Wu, 2009), geographic regions (Zhang and Kanbur, 2005; Hannum and Wang, 2006; Heckman, 2005) and sibling size (Li, Zhang, and Zhu (2008; Rosenzweig and Zhang, 2009) all contribute to educational inequality.

d. The data used in this study

In this study I follow Golley and Kong (2018)'s article which studies population with birth years from 1940 to 1989 based on the 2010 CFPS baseline survey. For consistency in comparisons, I apply both 2010 baseline and 2012~2016 follow-up surveys, and analyze the population born from 1940 to 1994 (people born in 1994 should finish their education except for those who are pursuing a master/doctoral degree, which accounts for little percentage in the total population and thus have no substantial effect on the result).

I choose the following circumstance variables: hukou[8] status (urban hukou=1, rural and other hukou=0); gender (male=1, female=0); father's education level (illiterate is the excluded category, 3 dummy variables: primary school, junior high, and senior high and above); whether parents are Communist Party member (=1 if either of parents is member); number of siblings (only child is the excluded category, 2 dummy variables: 1 or 2 siblings and 3 or more siblings); ethnicity (Han=1, minority=0); and residing provinces (Beijing is the excluded category). There are 17,250 valid data entries with all of the above information available.

I choose father's educational level as the representation of parental social status. Ferreira and Gignoux (2011) include both father and mother's education in their study, but In China it appears a high degree of marriage matching (Grier et al. 2016), which results in high correlation between parents' socioeconomic status, including educational background. Including both parent's education level will bias the estimated coefficient because of the correlation between variables (Knight et al. 2012). Likewise, variables like parents' occupation and income are not included.

[8] Hukou is a household registration system that differentiate households' activity region. In China, children with rural hukou can only go to rural schools.

Results

Table I.0. Similarities and differences between my analysis and Golley & Kong (2018)

My analysis	Golley and Kong (2018)
Similarities	
Gaps in average years of schooling: 1.6 years between male and female, 2 years for children whose parents with and without Communist Party membership, 1.7 years between Han Chinese and ethnic minorities, 3 years between only child and those with three or more siblings.	Gaps in average years of schooling: 1.4 years between male and female, 2.2 years for children whose parents with and without Communist Party membership, 2.3 years between Han Chinese and ethnic minorities, 3.8 years between only child and those with three or more siblings.
Below coefficients in OLS regression have positive signs and indicates individual with such circumstance variable may receive more years of schooling than baseline group in overall data. Urban hukou: 2 years more Male (until 1984): 1.6 years more Parents with Communist Party membership: 1 year more Father's education in 3 higher levels: 2, 3, 3.8 years more Effect of sibling size is inconclusive.	Below coefficients in OLS regression have positive signs and indicates individual with such circumstance variable may receive more years of schooling than baseline group in overall data. Urban hukou: 3 years more Male (until 1984): 1.4 years more Parents with Communist Party membership: 0.8 year more Father's education in 3 higher levels: 2, 2.2, 3.2 years more Effect of sibling size is inconclusive.
The relative measurement (direct and indirect) ranges from 22.4% (22.0%) in 1955-59 cohort to 42.1%(38.1%) in 1980-84 cohort.	IOR_D (IOR_I) ranging from a low of 25.3 (23.4) per cent for the 1960–64 cohort to a high of 42 (43.4) per cent for the 1981–85 cohort.
Top 3 contributors to educational inequality of opportunity are :father's education, urban hukou, and province.	Top 3 contributors to educational inequality of opportunity are :urban hukou, father's education, and province.
Differences	
I try to select the top 2 most important circumstance variables and form 8 types with sufficient observations in each type.	Golley & Kong divide sample into 192 types and leave 2/5 types with less than 1 observation.

a. Descriptive statistics

The average years of schooling shows an upward trend over the years whereas we can notice the gaps of year of schooling in different circumstances: 3.1 years between urban and rural hukou; 1.6 years between male and female; 2 years between parents with and without Communist party membership; 1.7 years between Han people and the minorities; 3 years between only child and those with 3 or more siblings.

From the summary statistic (Table I.1), the overall average years of schooling shows an upward trend over the years, with the illiterate decreasing from about half of the population to less than 10%, credited to high quality implementation of 9-year compulsory education system starting 1982 . Except for the cultural revolution period (1966-1976), the rate of people with low education attainment (primary school and junior high school) declines while more and more people are able to receive education from senior high school and above, with a notable jump of more than 10% from the year cohort 85-89 to 90-94, credited to the college enrollment expansion policy starting in 1999.

Table I.1. Descriptive statistics through different birth year cohort, for available provinces in China

Circumstance Variables	all	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94
#observation	17250	798	1408	1946	1835	2118	2534	2341	1591	1316	1365	1233
Education(%)												
Illiterate	27.4	46.6	46.0	46.1	36.9	21.4	26.2	23.2	15.7	10.0	7.2	7.2
Primary school	22.0	26.1	28.0	22.3	13.9	18.1	27.7	25.9	21.1	19.6	16.6	11.9
Junior high school	30.9	16.3	18.0	22.0	28.1	36.6	32.9	33.7	37.1	37.5	38.5	31.0
Senior high school and above	19.6	11.0	8.0	9.6	21.1	24.0	13.2	17.2	26.0	33.0	37.7	52.5
Average schooling (years)	6.7	4.5	4.3	4.5	6.0	7.4	6.4	6.9	8.2	9.2	9.7	11.4
Father's Education(%)												
Illiterate	49.3	77.8	72.2	69.9	62.7	58.5	50.5	37.6	28.7	21.4	16.7	15.3
Primary school	28.1	17.7	20.1	21.1	26.8	27.0	29.8	36.1	38.7	29.2	25.9	22.2
Junior high school	14.0	2.3	3.9	5.4	6.8	8.0	11.5	17.3	21.2	30.9	36.9	41.2
Senior high school and above	8.5	2.3	3.8	3.5	3.7	6.5	8.2	9.0	11.5	18.5	20.5	21.2
Father's average schooling (years)	4.0	1.5	2.1	2.2	2.7	3.2	3.9	4.9	5.7	6.8	7.4	7.8
Hukou (Urban=1, %)	48.2	44.7	47.5	46.5	46.7	45.4	43.9	49.5	55.1	52.8	52.7	57.0
Urban average schooling (years)	8.3	5.9	5.7	6.0	7.1	8.7	7.9	8.6	9.9	11.0	10.9	12.4
Rural average schooling (years)	5.2	3.3	3.1	3.3	5.0	6.3	5.1	5.3	6.0	7.2	8.2	10.1
Gender (Male=1, %)	48.7	55.6	52.6	51.1	50.1	47.0	47.4	46.4	47.8	46.9	47.5	51.0
Male average schooling (year)	7.5	5.7	5.3	5.7	7.4	7.8	7.1	7.7	8.8	9.5	9.8	11.5
Female average schooling (year)	5.9	2.9	3.3	3.3	4.5	6.4	5.7	6.2	7.6	8.9	9.5	11.4
Parents party member (yes=1, %)	15.2	5.8	8.5	14.0	17.2	19.6	17.8	17.7	15.6	14.7	11.1	11.6
Parents party member average schooling (year)	8.4	5.7	5.3	6.1	7.2	7.9	7.9	8.7	9.9	11.6	11.6	13.0
Parents non-party member average schooling (year)	6.4	4.4	4.2	4.3	5.7	7.1	6.0	6.5	7.9	8.8	9.4	11.2

Ethnicity (Han=1, %)	92.4	92.7	93.8	93.9	93.6	93.9	92.6	91.9	89.9	89.8	90.8	90.5
Han average schooling (year)	6.8	4.6	4.4	4.6	6.0	7.6	6.4	7.1	8.5	9.5	9.9	11.8
Minority average schooling (year)	5.1	3.3	3.1	3.7	4.9	5.7	5.3	4.7	5.3	6.1	7.6	7.5
Family size(%)												
Only child	7.3	8.6	7.2	4.9	5.1	3.6	2.6	3.4	9.7	19.3	20.1	26.0
One or two siblings	35.1	27.4	24.9	20.9	17.8	21.0	25.7	42.7	59.1	61.2	66.3	65.4
Three or more siblings	57.6	63.9	67.9	74.2	77.1	75.3	71.7	53.9	31.2	19.5	13.6	8.7
Only child average schooling (year)	8.8	3.8	4.2	5.2	5.6	6.8	5.8	7.4	10.4	11.9	11.9	13.4
One or two siblings average schooling (year)	7.7	4.3	4.0	4.5	6.5	7.5	7.5	7.8	8.5	9.1	9.4	11.2
Three or more siblings average schooling (year)	5.8	4.6	4.5	4.5	5.9	7.5	6.0	6.1	6.8	6.7	7.4	7.5

Source: CFPS 2010-2016

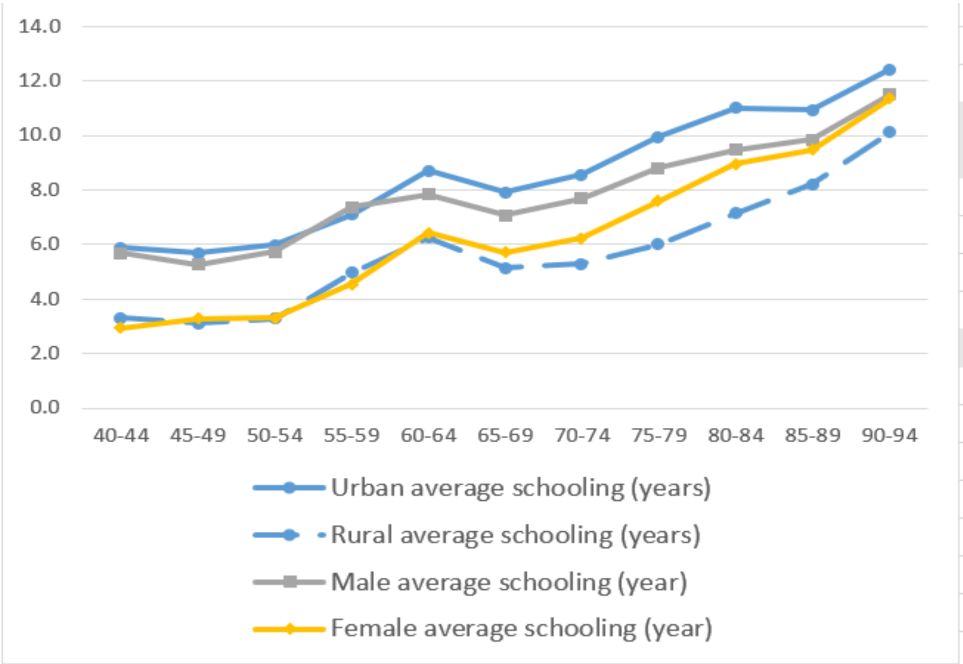


Figure I.3. Trends in average years of schooling – rural/urban and gender divides by birth year cohort, for available provinces in China
Source: CFPS 2010-2016

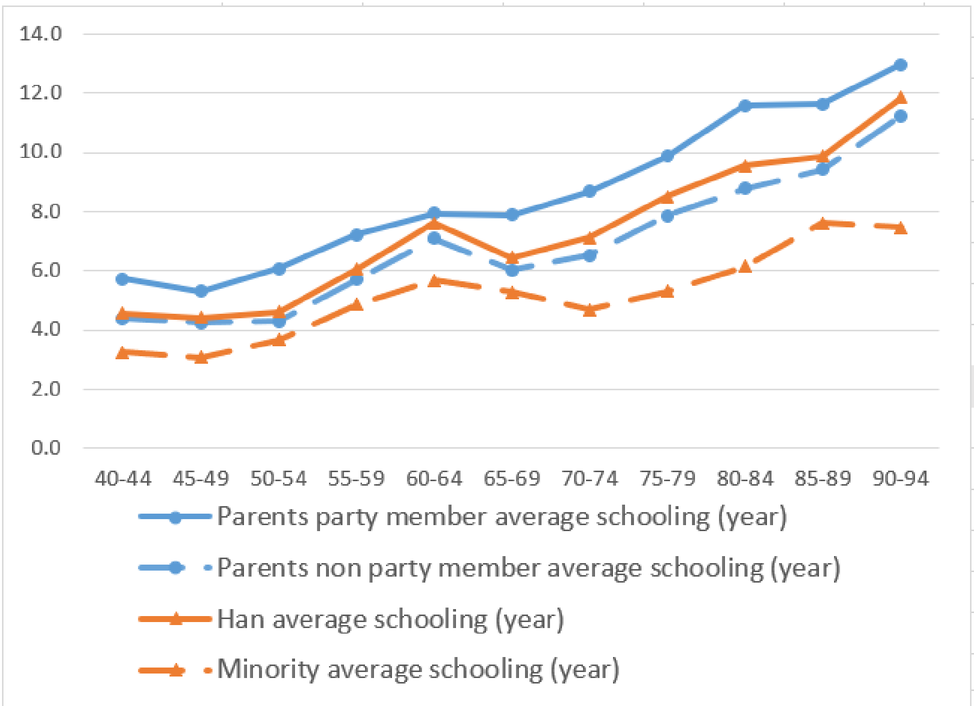


Figure I.4. Trends in average years of schooling – parents political party and ethnicity divides by birth year cohort, for available provinces in China
Source: CFPS 2010-2016

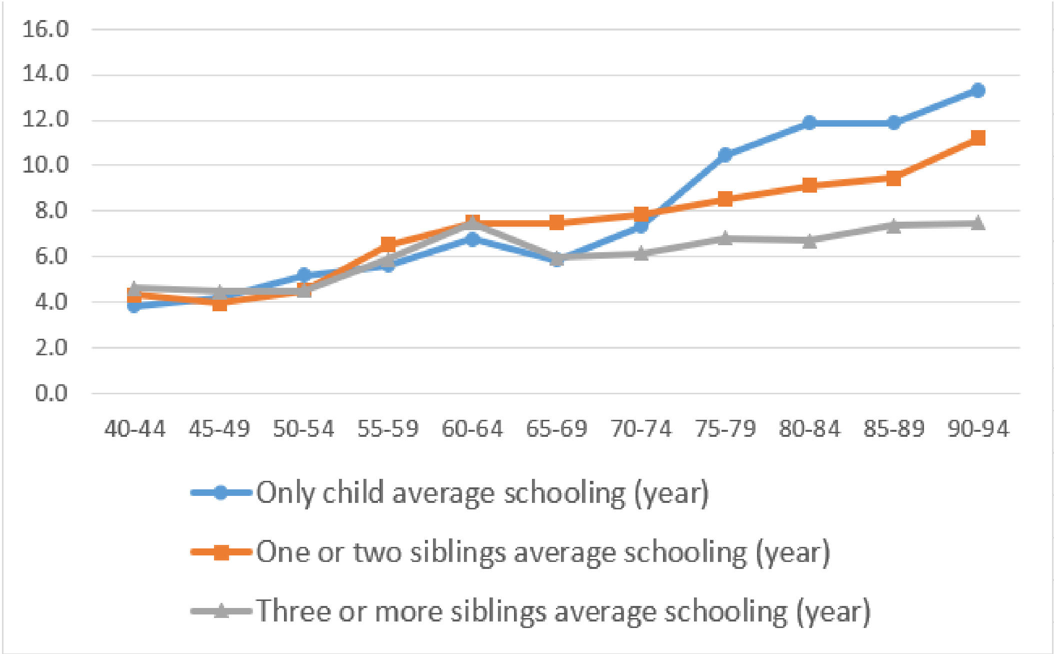


Figure I.5. Trends in average years of schooling – sibling size divides by birth year cohort, for available provinces in China.
Source: CFPS 2010-2016

Trends in average years of schooling by birth year cohort are presented by above figures. The urban-rural gap (Figure I.3) is relatively constant on and above 3 years. The gender gap (FigureI. 3) is constantly decreasing, from 3 years to almost 0. Traditionally, Chinese family value male children over girl children because of the culture norm (male is considered the true inheritor of the family), thus families allocate limited educational resources to boys. Because of the compulsory 9-year education policy and one-child policy, the gaps between boys and girls are disappearing. There are persist gaps (Figure I.4) between the Han ethnicity and the minorities, and between parents with or without the Communist Party membership. As for the relationship between sibling size and average year of schooling (Figure I.5), before the 1975-79 cohort, there is no clear relationship. Starting from the 75-79 birth cohort, children living in a family with less sibling size tend to enjoy more years of schooling. The one-child policy is initiated in the late 1970s and early 1980s, and it penalizes families give birth to addition children. Hence families with multiple children tend to have more restricted educational resources to allocated for each child. Becker and Lewis (1973) believe that parents will choose to have fewer children to invest in the human capital, in order to achieve the higher quality of offspring.

b. Regression results

The regression result is calculated based on equation (2) $y=\beta C+\varepsilon$ mentioned in Methods section to generate the parametrically standardized distributions. In this case y is years of schooling, C is a vector of circumstance variables including hukou status, gender, father's education level, whether parents are Communist Party member, number of siblings, ethnicity, and residing provinces as mentioned in section Data (d). Except for the geographic dummies, all coefficients have expected signs and mostly statistically significant.

Table I.2. OLS regression of years of schooling on circumstance variables by birth year cohort, for available provinces in China

Explanatory Variables	all	t	40-44	t	45-49	t	50-54	t
Hukou (Urban=1)	**2.046**	30.39	**1.816**	5.65	**2.215**	9.89	**2.120**	10.61
Gender (Male=1)	**1.601**	25.83	**3.112**	10.85	**2.093**	10.03	**2.630**	14.5
Parents party member (Yes=1)	**1.007**	11.47	1.052	1.72	0.556	1.48	**0.954**	3.58
Ethnicity (Han=1)	**0.488**	3.62	-0.602	-0.9	-0.582	-1.09	-0.0311	-0.07
Father-primary school (Yes=1)	**2.054**	27.53	**2.398**	6.33	**1.543**	5.76	**1.830**	8.01
Father-junior high (Yes=1)	**2.999**	30.81	**2.548**	2.64	**1.765**	3.27	**1.409**	3.46
Father-senior high and over (Yes=1)	**3.830**	31.98	0.721	0.74	**2.100**	3.81	**1.850**	3.7
One or two siblings (Yes=1)	**-0.288**	-2.26	0.547	0.98	0.309	0.7	-0.656	-1.44
Three or more siblings (Yes=1)	**-1.284**	10.26	0.235	0.45	0.462	1.13	-0.777	-1.83
Tiajing	-0.735	-1.43	**-8.968**	-3.08	-1.838	-1.08	**-4.123**	-2.85
Hebei	**-1.686**	-4.39	**-4.767**	-2.45	-2.311	-1.68	-1.906	-1.79
Shanxi	**-1.441**	-3.67	**-5.480**	-2.78	-2.796	-1.95	**-3.262**	-2.98
Neimenggu	-2.005	-0.97	-		-3.458	-0.84	-	
Liaoning	**-1.268**	-3.37	-3.412	-1.86	-1.387	-1.03	-1.527	-1.47
Jilin	**-1.791**	-4.25	-3.405	-1.6	-2.847	-1.82	**-2.641**	-2.2
Heilongjiang	**-1.731**	-4.33	**-4.960**	-2.37	-3.274	-2.28	**-2.402**	-2.12
Shanghai	**-1.054**	-2.77	**-4.352**	-2.37	-1.539	-1.15	**-2.125**	-2.04
Zhejiang	**-2.371**	-5.61	**-5.450**	-2.78	**-4.834**	-3.34	**-5.076**	-4.26
Jiangsu	**-1.473**	-3.42	**-5.929**	-2.67	-2.478	-1.59	**-4.498**	-3.46
Anhui	**-3.583**	-8.34	**-7.913**	-3.91	**-5.326**	-3.53	**-6.384**	-5.43
Fujian	**-2.934**	-5.59	**-8.291**	-3.61	**-6.292**	-3.61	**-3.895**	-2.66
Jinagxi	**-2.828**	-6.67	**-5.457**	-2.68	**-3.249**	-2.15	**-3.718**	-2.98
Shandong	**-2.164**	-5.59	**-4.571**	-2.36	**-3.280**	-2.38	**-3.945**	-3.72
Henan	**-1.728**	-4.61	**-5.530**	-2.99	-2.544	-1.9	**-2.518**	-2.42

Table I.2. continue: OLS regression of years of schooling on circumstance variables by birth year cohort, for available provinces in China

Explanatory Variables	all	t	40-44	t	45-49	t	50-54	t
Hubei	**-1.570**	-3.57	**-4.932**	-2.32	**-3.497**	-2.29	**-2.746**	-2.18
Hunan	**-0.877**	-2.16	**-4.234**	-2.17	-1.595	-1.12	-1.596	-1.44
Guangdong	**-2.252**	-5.93	**-6.443**	-3.44	**-2.904**	-2.14	**-3.045**	-2.85
Guangxi	**-1.865**	-4.35	**-5.986**	-2.75	**-3.968**	-2.51	**-3.619**	-2.76
Hainan	**-5.922**	-2.87	-		-3.478	-0.85	-	
Chongqing	**-3.102**	-6.41	**-5.941**	-3.01	**-3.862**	-2.55	**-4.018**	-3.02
Sichuan	**-3.137**	-8.03	**-6.115**	-3.18	**-3.965**	-2.87	**-4.207**	-3.87
Guizhou	**-3.836**	-9.19	**-8.303**	-4.21	**-5.773**	-3.93	**-4.796**	-4.05
Yunnan	**-2.957**	-7.28	**-7.846**	-3.42	**-5.261**	-3.44	**-3.087**	-2.49
Xizang	-		-		-		-	
Shan'xi	**-1.572**	-3.78	**-5.024**	-2.48	-2.309	-1.58	**-2.531**	-2.16
Gansu	**-3.085**	-8.2	**-6.563**	-3.52	**-3.462**	-2.56	**-3.794**	-3.61
Qinghai	-1.912	-0.66	-		-		-	
Ningxia	-		-		-		-	
Xinjiang	**-4.669**	-2.75	-		-		-	
Constant	**5.909**	14.54	**7.042**	3.55	**4.708**	3.24	**5.267**	4.42
Observations	17,250		798		1,408		1,946	
R-squared	0.291		0.311		0.252		0.268	

Table I.2. continue: OLS regression of years of schooling on circumstance variables by birth year cohort, for available provinces in China

Explanatory Variables	55-59	t	60-64	t	65-69	t	70-74	t
Hukou (Urban=1)	**1.533**	6.75	**1.861**	9.88	**1.988**	12.11	**2.310**	13.6
Gender (Male=1)	**2.899**	13.97	**2.039**	11.79	**1.447**	9.58	**1.531**	9.7
Parents party member (Yes=1)	**0.930**	3.31	**0.931**	4.19	**1.019**	5.08	**1.199**	5.72
Ethnicity (Han=1)	0.286	0.58	**0.860**	2.19	0.355	1.1	**1.257**	3.76
Father-primary school (Yes=1)	**1.289**	5.31	**1.510**	7.4	**1.769**	9.93	**1.424**	7.56
Father-junior high (Yes=1)	**1.197**	2.84	**1.490**	4.53	**1.854**	7.37	**2.149**	9.04
Father-senior high and over (Yes=1)	**1.556**	2.78	**2.195**	6	**2.753**	9.34	**3.441**	11.31
One or two siblings (Yes=1)	0.778	1.49	0.756	1.54	**1.131**	2.31	0.533	1.2
Three or more siblings (Yes=1)	0.515	1.08	0.507	1.09	0.528	1.11	-0.166	-0.38
Tiajing	-2.911	-1.8	-0.578	-0.36	0.263	0.18	1.320	0.89
Hebei	**-3.972**	-3.27	-2.192	-1.88	-1.098	-1.04	-0.457	-0.4
Shanxi	**-3.385**	-2.73	-1.645	-1.41	0.0621	0.06	-0.486	-0.42
Neimenggu	-		-		-0.160	-0.06	-	
Liaoning	**-3.800**	-3.19	**-2.516**	-2.22	-0.211	-0.2	0.776	0.69
Jilin	**-3.960**	-2.99	**-2.630**	-2.14	-0.394	-0.34	-0.617	-0.51
Heilongjiang	**-4.022**	-3.17	-2.088	-1.75	-0.217	-0.2	0.274	0.23
Shanghai	**-2.609**	-2.18	-1.619	-1.41	0.445	0.41	0.411	0.36
Zhejiang	**-5.299**	-3.93	**-3.149**	-2.44	-0.449	-0.38	0.683	0.55
Jiangsu	**-5.594**	-4.05	-1.824	-1.47	-1.525	-1.33	-0.646	-0.51
Anhui	**-5.805**	-4.04	**-3.768**	-2.93	-2.190	-1.95	-2.207	-1.79
Fujian	**-3.576**	-1.98	**-4.908**	-2.85	**-4.681**	-2.79	**-3.346**	-2.06
Jinagxi	**-5.480**	-4.16	**-4.236**	-3.4	**-2.268**	-1.98	-2.089	-1.72
Shandong	**-3.771**	-3.09	**-3.117**	-2.68	-1.190	-1.12	-0.935	-0.81
Henan	**-2.872**	-2.4	**-2.475**	-2.19	-0.751	-0.72	-0.909	-0.81

Table I.2. continue: OLS regression of years of schooling on circumstance variables by birth year cohort, for available provinces in China								
Explanatory Variables	55-59	t	60-64	t	65-69	t	70-74	t
Hubei	**-3.463**	-2.37	-1.331	-1.01	-0.662	-0.57	-0.326	-0.26
Hunan	**-2.925**	-2.22	-2.067	-1.71	0.843	0.76	-0.0793	-0.07
Guangdong	**-4.506**	-3.72	**-3.269**	-2.84	-1.943	-1.87	-0.913	-0.81
Guangxi	-2.689	-1.88	**-2.736**	-2.13	-0.531	-0.47	-0.299	-0.25
Hainan	-		-		-		**-8.177**	-2.83
Chongqing	**-5.291**	-3.21	-2.724	-1.93	-1.805	-1.26	-1.783	-1.31
Sichuan	**-6.487**	-5.12	**-3.505**	-3.02	-0.906	-0.84	-1.959	-1.69
Guizhou	**-5.635**	-4.15	**-4.368**	-3.35	**-3.583**	-3.17	**-2.532**	-2.12
Yunnan	**-5.265**	-3.95	**-5.024**	-4.12	**-2.326**	-2.14	-1.564	-1.34
Xizang	-		-		-		-	
Shan'xi	**-3.721**	-2.72	-2.309	-1.87	-0.367	-0.32	0.213	0.18
Gansu	**-5.957**	-4.96	**-3.460**	-3.05	**-2.528**	-2.44	**-2.208**	-1.97
Qinghai	-		-		-		-	
Ningxia	-				-		-	
Xinjiang	-				-2.150	-0.55	**-7.283**	-2.52
Constant	**6.498**	4.88	**6,141**	4.86	**3.831**	3.27	**3.215**	2.62
Observations	1,835		2116		2,534		2,341	
R-squared	0.220		0.227		0.274		0.328	

Table I.2. continue: OLS regression of years of schooling on circumstance variables by birth year cohort, for available provinces in China

Explanatory Variables	75-79	t	80-84	t	85-89	t	90-94	t
Hukou (Urban=1)	**2.533**	12.75	**2.119**	9.36	**1.406**	6.81	**0.81**	3.2
Gender (Male=1)	**0.976**	5.36	0.332	1.66	-0.0289	-0.15	-0.0370	-0.16
Parents party member (Yes=1)	**0.734**	2.85	**1.285**	4.43	**1.012**	3.23	0.108	0.28
Ethnicity (Han=1)	**1.269**	3.55	**1.312**	3.4	0.645	1.61	**1.835**	3.73
Father-primary school (Yes=1)	**1.474**	6.4	**1.488**	5.08	**1.794**	5.82	**1.784**	4.49
Father-junior high (Yes=1)	**2.478**	9.04	**1.982**	6.59	**2.775**	9.31	**2.260**	6.02
Father-senior high and over (Yes=1)	**3.766**	11.07	**2.702**	7.81	**3.802**	11.26	**3.973**	9.23
One or two siblings (Yes=1)	-0.536	-1.61	**-1.076**	-3.68	**-1.173**	-4.23	**-1.342**	-4.34
Three or more siblings (Yes=1)	**-1.121**	-3.07	**-2.507**	-6.82	**-2.352**	-6.19	**-3.330**	-6.48
Tiajing	1.805	1.41	1.071	0.8	-0.675	-0.42	-1.655	-1.1
Hebei	-0.0127	-0.01	-1.424	-1.29	**-2.147**	-2.24	**-2.405**	-2.34
Shanxi	-0.806	-0.79	-0.534	-0.46	-1.616	-1.64	**-3.506**	-3.42
Neimenggu	-		-		-0.595	-0.17	1.844	0.75
Liaoning	0.508	0.52	-1.612	-1.5	-1.752	-1.82	**-2.446**	-2.42
Jilin	-1.139	-1.04	-2.006	-1.65	-2.217	-1.93	-1.610	-1.24
Heilongjiang	-1.570	-1.54	-2.239	-1.92	**-2.346**	-2.11	**-2.726**	-2.35
Shanghai	0.953	0.95	0.431	0.4	0.496	0.5	**-2.668**	-2.56
Zhejiang	0.138	0.12	-0.656	-0.52	-1.671	-1.56	**-2.454**	-2.31
Jiangsu	0.997	0.87	0.600	0.5	-0.476	-0.43	-1.857	-1.55
Anhui	-1.499	-1.32	-0.955	-0.7	-2.131	-1.69	**-2.819**	-2.4
Fujian	0.437	0.33	-0.545	-0.4	-1.992	-1.5	-3.063	-1.9
Jinagxi	-0.730	-0.65	**-2.696**	-2.14	-1.325	-1.14	**-3.028**	-2.25
Shandong	0.0275	0.03	-1.191	-1.05	-1.527	-1.5	-1.636	-1.54
Henan	0.265	0.27	-1.112	-1.03	**-2.047**	-2.18	**-3.290**	-3.42

Table I.2. continue: OLS regression of years of schooling on circumstance variables by birth year cohort, for available provinces in China

Explanatory Variables	75-79	t	80-84	t	85-89	t	90-94	t
Hubei	0.484	0.43	-0.491	-0.37	-2.173	-1.68	-0.385	-0.29
Hunan	1.013	0.96	0.630	0.52	-1.208	-1.07	**-3.106**	-2.6
Guangdong	-0.418	-0.43	-1.209	-1.11	-1.655	-1.72	**-2.044**	-2.13
Guangxi	-0.0417	-0.04	**-2.563**	-2.16	**-2.734**	-2.16	**-4.678**	-3.65
Hainan	-2.053	-0.56	-		-		-	
Chongqing	-0.310	-0.19	**-4.954**	-2.61	-1.933	-1.32	-3.529	-1.42
Sichuan	**-2.192**	-2.15	**-3.068**	-2.63	**-2.968**	-2.98	**-4.701**	-4.47
Guizhou	**-2.585**	-2.38	-1.355	-1.09	-1.240	-1.11	-2.256	-1.8
Yunnan	-1.666	-1.64	-1.820	-1.54	**-2.759**	-2.67	**-3.936**	-3.52
Xizang	-		-		-		-3.331	-1.12
Shan'xi	-1.187	-1.13	-1.280	-0.99	**-2.279**	-2.09	**-3.263**	-2.84
Gansu	**-2.224**	-2.3	**-2.422**	-2.23	**-2.765**	-2.9	**-2.737**	-2.85
Qinghai	-		-		-3.084	-1.17	-	
Ningxia	-		-		-		-2.333	-0.78
Xinjiang	-1.612	-0.71	-		-		0.382	0.17
Constant	**4.786**	4.51	**7.480**	6.58	**8.941**	8.64	**11.04**	10.03
Observations	1,591		1,316		1,365		1,233	
R-squared	0.399		0.381		0.316		0.280	

Notes: Robust t statistics available; variables with bold text have 5% statistical significance level or better. Due to violation of OLS assumptions in model test, the significance level may be overestimated.

Resource: CFPS 2010-2016 and author's calculation based on $y = \beta C + \varepsilon$

The coefficients are interpreted as *additional years of schooling compared to the baseline group* with exclude dummy circumstance variables, which is the only child of the family, lives in Beijing, and has an illiterate father. We can observe that urban hukou enjoys additional 2 years of schooling than rural hukou, comparing to 3.1 years in the summary data. The coefficient of urban hukou is positive and statistically significant in all year cohorts. Male also appears to enjoy additional 1.6 years of schooling in the overall sample, which consists with the preliminary data. The gender coefficient becomes less important after the 1980-84 cohort because of the one-child policy. Children whose parents are Communist party member tend to receive more years of schooling. Coefficients of Han ethnicity shows both signs, but only positive ones are statistically significant. Fathers with higher education increase the possibility of children receiving more years of schooling, proven by the all positive and statistically significant coefficients, with respectively additional 2, 3, and 3.8 years for primary school, junior high school and senior high school and above. Coefficients for having 1 or 2 and 3 or more siblings appear in both signs, while the statistically significant ones after 1980-84 cohort and the overall sample illustrate that being the only child enjoys more years of schooling. Most geographic dummies show statistically significant and negative coefficients (except for Shanghai, Zhejiang and Jiangsu in several cohort, though statistically insignificant), which suggests children in Beijing receive the longest time of education.

We can also infer certain aspect of the intergenerational mobility. Children whose father has higher education tend to have more education, though the coefficient differs in different year cohort. For example, for fathers with senior high school and higher education, we can observe that the stickiness of intergenerational mobility is much higher in recent cohorts (the 1985-94 cohort) than the 1955-60 cohort. People born in the later cohort are influenced largely by the culture revolution (1966-76), and hence are less influenced by their fathers' educational attainment.

As mentioned in Methods section, we need to test the OLS model due to various reasons, including the omitted and unobserved circumstances, which could lead to biased coefficients and thus do not necessarily indicate causal relationship. Normality test (Figure I.A1 and A2 in appendix) indicates the residual is not normally distributed, thus the standard errors of OLS estimates is not perfectly reliable, which means the confidence intervals would be too narrow. However, normality is not required in order to obtain unbiased estimates of the regression coefficients. Multicollinearity test (Table I.A1 in appendix) proves most circumstance factors are not linearly related to each other, except for a few geographic dummy variables. As a rule of thumb, a variable whose VIF values are greater than 10 may be problematic. 4 provincial variables with VIF value a little bit over than 10 could imply similar educational resource endowment among provinces but should not affect the overall model. Model Specification test (Figure I.A3 in appendix) has failed to reject the hypothesis that the model is specified correctly. Therefore, it seems that we don't have a specification error. Breusch-Pagan/Cook-Weisberg test for heteroskedasticity (in Appendix) shows a very small p value and we have to accept the alternative hypothesis that the variance is not homogenous. Cross-sectional studies are more likely to have heteroscedasticity because they often have values differ in a big range. While heteroscedasticity does not cause bias in the coefficient estimates, it does make them less precise and further from the correct population value. Heteroscedasticity increases the variance of the coefficient estimates and tends to produce p-values that are smaller than the actual value, and in turn a statistically important coefficient could be not statistically important. Violation of the OLS assumptions will result in underestimated confidence intervals and underestimated p-value, and thus the statistical significance level of the independent variable may also be overestimated. Hence the statistical significance of the coefficients should be interpreted with caution. Nevertheless, the main goal of the article is to explore to what extent is the inequality in years of schooling attributable to circumstance variables, which will be shown in the next section.

c. Inequality of opportunity

First, I replicated Golley and Kong (2018)'s article on CFPS 2010&2012 data and get very similar results below. (Tables I.3 and Figure I.6. The regression results and partial contribution to relative inequality of opportunity can be found in Appendix.) We find both direct and indirect methods show close measurement of inequality in absolute and relative terms. The inequality of opportunity ranges from 26.5%(26.0%) in 1960-64 cohort to 41.9%(42.4%) in 1980-84 cohort, in total inequality of years of schooling.

Table I.3: Inequality of opportunity in years of schooling by birth year cohort, for available provinces in China

Birth cohort	All	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89
Total outcome inequality											
GE(2)	0.292	0.626	0.600	0.532	0.359	0.193	0.264	0.250	0.291	0.124	0.097
Gini	0.426	0.596	0.583	0.561	0.470	0.342	0.402	0.393	0.426	0.275	0.242
Direct											
IOp (within group)	0.199	0.431	0.426	0.364	0.250	0.142	0.194	0.169	0.198	0.072	0.062
Absolute: IOA_D (between	0.093	0.195	0.170	0.168	0.109	0.051	0.070	0.080	0.093	0.052	0.034
Relative: IOR_D	0.318	0.312	0.283	0.316	0.302	0.265	0.265	0.322	0.319	0.419	0.355
Indirect											
GE(2) y average	0.197	0.424	0.424	0.359	0.248	0.143	0.186	0.165	0.197	0.072	0.065
IOA_I	0.095	0.202	0.176	0.173	0.111	0.050	0.078	0.085	0.094	0.053	0.032
IOR_I	0.327	0.323	0.293	0.325	0.309	0.260	0.297	0.341	0.325	0.424	0.331

Source: CFPS 2010 & 2012 and author's calculation based on regression results in Appendix and method described in section Methods.

	All	By birth cohort									
		40–44	45–49	50–54	55–59	60–64	65–69	70–74	75–79	80–84	85–89
Total (outcome) inequality											
GE(2)	0.207	0.624	0.585	0.538	0.313	0.158	0.198	0.185	0.134	0.105	0.085
Gini	0.355	0.588	0.572	0.557	0.435	0.302	0.341	0.331	0.280	0.248	0.224
Inequality of opportunity											
Absolute: IOA_D	0.082	0.194	0.143	0.176	0.086	0.040	0.056	0.062	0.052	0.045	0.029
IOA_I	0.069	0.253	0.225	0.215	0.088	0.037	0.048	0.057	0.053	0.045	0.030
Relative: IOR_D	0.393	0.311	0.245	0.327	0.276	0.253	0.284	0.338	0.388	0.431	0.338
IOR_I	0.334	0.406	0.384	0.400	0.280	0.234	0.244	0.309	0.394	0.434	0.355

Figure I.6: Inequality of opportunity in years of schooling by birth year cohort from Golley and Kong's paper (2018)

Source: Golley and Kong (2018)

Second, I repeat the same method on with updated data (Table I.4 and Figure I.7). We can observe slightly smaller inequality measurements (both direct and indirect terms) from the updated 2010-2016 data than the 2010 baseline data. Similarly, both direct and indirect methods are very close in absolute and relative terms. The relative measurement ranges from 22.4%(22.0%) in 1955-59 cohort to 42.1%(38.1%) in 1980-84 cohort. The measurements of total inequality, GE(2) and Gini[9] index both have declining trends except the sudden drop in the 1960-64 cohort when most influenced by the infamous cultural revolution, which targeted to eliminate the "privileged" with high socioeconomic status and pursued Egalitarianism. The severe economic recession caused by the policy suggests that judging inequality outcomes only on total inequality or absolute term of inequality is unreliable and misleading.

The relative term of inequality better reflects concepts of social justice. The inequality of opportunity is widening since the 1950-54 cohort and starts to mitigate from the 1980-84 cohort. The increase of inequality of opportunity during the cultural revolution periods further proves the failure of the policy. On the other hand, the declining trend since 1980 suggests the 9-year compulsory education policy which enacts in 1982 plays an important role in improving social justice.

[9] We use Gini index as comparison here. Gini index is a feasible measurement of total inequality, but it is not decomposable.

Table I.4: Inequality of opportunity in years of schooling by birth year cohort, for available provinces in China

By birth cohort	All	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94
Total outcome inequality												
GE(2)	0.257	0.542	0.514	0.512	0.344	0.181	0.238	0.219	0.154	0.116	0.092	0.082
Gini	0.395	0.558	0.544	0.545	0.455	0.325	0.376	0.361	0.301	0.261	0.230	0.223
Direct												
IOp (within group)	0.182	0.381	0.384	0.377	0.267	0.135	0.176	0.144	0.093	0.067	0.059	0.054
Absolute: IOA_D (between group)	0.075	0.162	0.130	0.135	0.077	0.047	0.063	0.074	0.061	0.049	0.034	0.028
Relative: IOR_D	0.292	0.298	0.253	0.263	0.224	0.257	0.263	0.339	0.398	0.421	0.364	0.338
Indirect												
GE(2) y average	0.183	0.374	0.384	0.375	0.268	0.140	0.173	0.147	0.093	0.072	0.063	0.059
IOA_I	0.075	0.169	0.130	0.137	0.076	0.041	0.065	0.072	0.062	0.044	0.029	0.023
IOR_I	0.291	0.311	0.252	0.268	0.220	0.227	0.274	0.328	0.399	0.381	0.316	0.279

Source: CFPS 2010-2016 and author's calculation based on regression results in Table 2 and method described in section Methods.

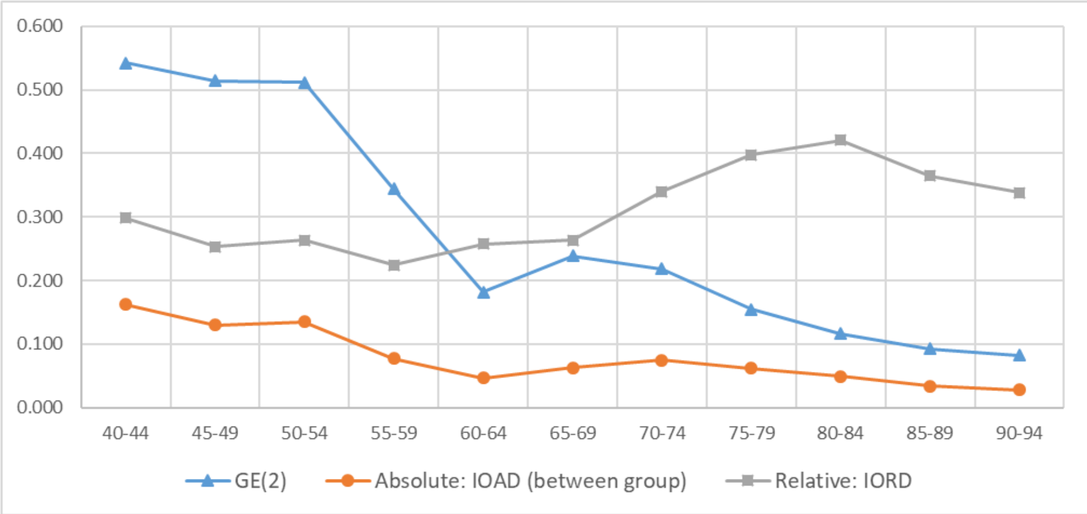


Figure I.7(a): Outcome inequality and non-parametric inequality of opportunity (absolute and relative) by birth year cohort from Table 4

Notes: GE(2) is half the coefficient of variation. IOA_D and IOR_D are direct absolute and relative measures of inequality of opportunity in Table 4

Source: CFPS 2010-2016 and author's calculation

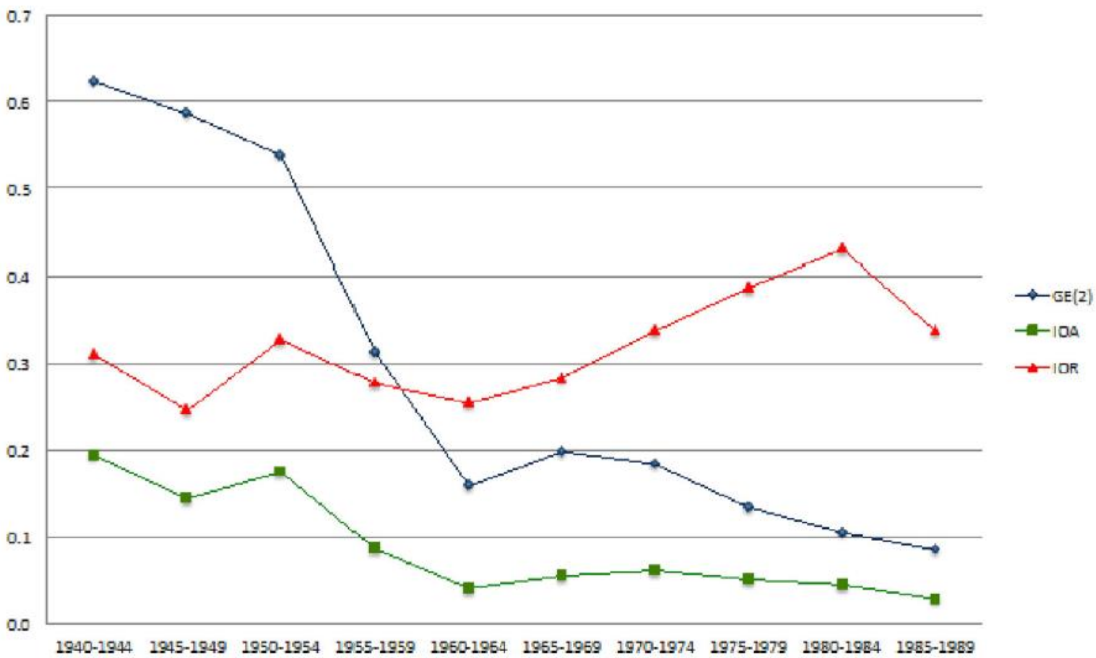


Figure I.7(b): Outcome inequality and non-parametric inequality of opportunity (absolute and relative) by birth year cohort from Golley and Kong's paper (2018)

Source: Golley and Kong (2018)

Partial contribution to relative inequality of opportunity (IORp) is calculated to rank the importance of each circumstance factor (in Table I.5). For example, to calculate the partial contribution of gender, I assign every individual in the sample with the average value for the variable gender, re-run the regression to generate the predicted years of schooling and calculate the absolute and relative term of inequality of the new distribution. Lower relative inequality in the new distribution with the specific equalized circumstance variable indicates that variable accounts for more responsibility in causing inequality of opportunity. For example, in Row 6 Column 2 of Table I.5, removing gender variable will cause reduction in inequality of opportunity, from 29.2% in Table I.4 to 26.4%.

The circumstance variables are ranked by their contribution to inequality of opportunity. Provinces, hukou, and father's education remain in top 3 with few exceptions, while the influence of gender becomes less important. The rank of ethnicity stays low, indicating relatively equal opportunity among the Han majority and ethnic minorities, considering the minorities usually reside in less developed areas with lower socioeconomic status. It suggests the policies aiming the ethnic minorities are effective. The increasing rank of sibling size implies the effectiveness of one-child policy. The high rank of father's education displays the existence of intergenerational transition of educational attainment. The two biggest contributors, hukou and province, agree on Golley and Kong's conclusion that where you live do matter. The hukou system worth special attention because it is unique in China and has huge influence.

Table I.5: Ranking of circumstance variables' partial contribution to relative inequality of opportunity by birth year cohort, for available provinces in China

Partial contribution	all	iorp	40-44	iorp	45-49	iorp	50-54	iorp	55-59	iorp	60-64	iorp	65-69	iorp	70-74	iorp	75-79	iorp	80-84	iorp	85-89	iorp	90-94	iorp
Urban hukou	2	0.202	4	0.242	2	0.168	2	0.187	3	0.178	1	0.154	3	0.183	2	0.212	1	0.237	1	0.232	3	0.234	2	0.190
Father's education	1	0.146	3	0.239	3	0.197	4	0.212	4	0.192	3	0.166	2	0.176	1	0.210	2	0.243	3	0.265	1	0.154	1	0.145
Province	3	0.228	1	0.198	1	0.159	1	0.185	2	0.138	2	0.161	1	0.176	3	0.219	3	0.248	2	0.255	2	0.225	3	0.196
Parents party member	6	0.274	5	0.307	5	0.249	5	0.254	5	0.209	5	0.211	5	0.253	4	0.303	7	0.384	5	0.346	5	0.297	6	0.278
Male	5	0.264	2	0.227	4	0.202	3	0.199	1	0.138	4	0.174	4	0.250	5	0.303	6	0.382	7	0.378	7	0.317	7	0.280
Number of siblings	4	0.256	6	0.313	6	0.251	6	0.267	6	0.218	7	0.223	6	0.259	6	0.307	4	0.369	4	0.285	4	0.240	5	0.245
Han	7	0.286	7	0.318	7	0.258	7	0.268	7	0.219	6	0.220	7	0.271	7	0.312	5	0.371	6	0.348	6	0.304	4	0.228

Note: Circumstance variables are ranked by their contribution to inequality of opportunity. Partial contribution to relative inequality of opportunity (IORp) is calculated by assigning variable mean to the group.

Source: CFPS 2010-2016 and author's calculation based on regression results above and method described in section Methods.

Interestingly, the educational inequality of opportunity in China reflects the rapid social and political changes in China since 1940s. When People's Republic of China was established in 1945, the goal of the phase was to reduce illiteracy, and universal education popularized rapidly. In the First Five Year Plan (1953~1957), educational resources were devoted to train skilled workers. During 1945~1957, the inequality of opportunity was high with a low overall education level. Great Leap Forward in 1958 shifted education agenda with political agenda to "left", and enrollment into primary level of education grew rapidly while curriculums showed more political tendency. During the decade of Cultural Revolution (1966-1976) policies shifted to extreme left and universities were closed. Mass education was promoted while elitism was disdained. During this period, the economy experienced a recession and educational inequality of opportunity showed unexpected increase. National entry test to universities (Gaokao) was not resumed until 1977. Along with reform and opening policy starting from 1978, education goal shifted to high quality education. The Compulsory Education Law in 1986 helped students in rural area to adopt primary and junior high school level of education. Since then, the overall educational level experiences constant increase while the inequality keeps narrowing down. The One-Child policy announced in early 1980s reduce the importance of gender's role playing in educational attainment. The college enrollment expansion in 1999 focused in urban area largely increase the population portion with higher than senior high school education. Emran and Sun (2011) claim such policy coincides with the fiscal decentralization which shift the financial burden on lower-level of government agencies and in turn intensify the urban-rural and provincial inequalities.

d. Alternative groups

Insufficient number of observations under each subgroup is a general problem researchers encounter using Roemer's type when measuring inequality. In this study we use 6 circumstance variables forming 192 types, which leaves more than 2/5 of groups with less than 1 observation. Ferreira and Gignoux (2011) claim that too few observations significantly impair the precision of the estimation.

Therefore, I choose the most important contributors (in Table I.5): hukou and father's education forming 8 types with sufficient observations in each group. The total inequality and parametric measurement of inequality are unaffected, while the non-parametric estimation of inequality decreases in value (comparing Table I.6 to Table I.4) but remains the same trend.

In reality, population can be divided according to many other circumstances. Ferreira and Gignoux (2011) claim omitting those circumstances and restricting the number of types because of limited data entry potentially leads to underestimation of inequality, and our result proves their point. Hence, the parametric method which overcomes above problems is useful in terms of reference, and we believe the reasonable range of inequality of opportunity is from 25% to 40%.

Table I.6: Inequality of opportunity in years of schooling with alternative group division by birth year cohort, for available provinces in China

By birth cohort	All	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94
Total outcome inequality												
GE(2)	0.257	0.542	0.514	0.512	0.344	0.181	0.238	0.219	0.154	0.116	0.092	0.082
Gini	0.395	0.558	0.544	0.545	0.455	0.325	0.376	0.361	0.301	0.261	0.230	0.223
Direct (alternative groups)												
IOp (within group)	**0.201**	**0.459**	**0.443**	**0.444**	**0.320**	**0.158**	**0.194**	**0.168**	**0.109**	**0.084**	**0.070**	**0.067**
Absolute: IOA_D (between group)	**0.057**	**0.084**	**0.071**	**0.068**	**0.024**	**0.023**	**0.044**	**0.050**	**0.046**	**0.032**	**0.022**	**0.015**
Relative: IOR_D	**0.220**	**0.155**	**0.137**	**0.133**	**0.071**	**0.128**	**0.186**	**0.230**	**0.296**	**0.274**	**0.239**	**0.185**
Indirect												
GE(2) y average	0.183	0.374	0.384	0.375	0.268	0.140	0.173	0.147	0.093	0.072	0.063	0.059
IOA_I	0.075	0.169	0.130	0.137	0.076	0.041	0.065	0.072	0.062	0.044	0.029	0.023
IOR_I	0.291	0.311	0.252	0.268	0.220	0.227	0.274	0.328	0.399	0.381	0.316	0.279

Source: CFPS 2010 - 2016 and author's calculation based on regression results in Table 2 and method described in section Methods.

Conclusion and discussion

In this article we find that as the average years of schooling increased steadily, the lower-bound inequality of opportunity constitutes 29% of the educational outcome inequality from CFPS data on population born from 1940 to 1994. The inequality of opportunity was increasing from the 1960s, passing through the Cultural Revolution, the reform and opening policy, one child policy, and the trend is declining since 1985. The most influential circumstances include hukou, father's education and province of residence. Urban hukou enjoy 3 more years of education controlling for other factors, and this result may result in rethinking the rationality of the hukou system.

The reasons for educational resource inequality are complex and comprise both unequal allocation of natural resources and human capital. In less developed areas, geographic limitations restrict access to resources. For example, in some extreme underdeveloped areas, students need to climb an 800-meter cliff in order to get to school[10]. The Chinese government has spent billions on improving the infrastructures and very few children need to face the cliff-climbing situations nowadays. In many rural villages, school building are the most safe and modern architectures, and classrooms have little visible differences with those in urban areas. However, rural students are still disadvantaged if we take teachers into consideration. Because of salary, transportation, and living conditions, most of the high-quality teachers stay in urban area. One attempt to mitigate this inequality is live-streaming teaching[11], that experienced teachers in famous schools

[10] Lu, S. (2016, May 27). China: Students make terrifying school run. CNN NEWS WORLD Retrieved from https://www.cnn.com/2016/05/27/europe/china-children-education-climb/index.html

[11] FENG, J (2018, December 13). Live-Streaming Offers Possible Solution to Education Inequality in China. Home Society & Culture Society.

live teach while in less developed areas, local teachers assist students to learn. Though the results of live teaching in China are as yet unproven, the approach offers a partial solution to teacher shortages in less developed areas. In addition, students from rural areas who succeed in their higher education probably choose to stay in urban areas and leave the rural area still deficient in human capital, forming a vicious cycle. Urban-rural education inequality is embedded in China along with the left-behind children problem. Parents coming from poor areas cannot support their family at their hometowns, and they migrate as temporary workers between their hometowns and large cities. Though parents' presence is proven to be critical to children's development (Thompson, 2014), China's hukou policy restricts them from obtaining urban hukou (legal residential rights) and hampers rural children from entering urban schools and living with their parents who work in urban areas. Children are usually left behind in the hometown with lower quality education, taken care of by grandparents, who are usually less educated. Left-behind children tend to leave school early because their parents can't give them support in mental, economic and social relationship skills. Lack of study motivation and clear career plans cause left-behind children more frequently leave school early and work to support their family. Su's study (2015) based on CHNS database finds that male students are more likely to drop out of school than female students because the market gives higher expected salary to male workers.

Many people did believe in the slogan for the compulsory education policy: "Education

---

Retrieved from https://supchina.com/2018/12/13/live-streaming-offers-possible-solution-to-education-inequality-in-china/

Changes the Destiny", because in the past half century, the social mobility in China was tremendous and many people achieved upward mobility taking the advantages of college enrollment expansion and substantial economic growth (Golley and Kong, 2013). However, more people are now questioning about the usefulness of education because of the emerging class solidification. Zhou and Xu's study (2017) report rural-urban children suffer lower marginal positive effect on their schooling outcomes result from their parents' educational attainment than urban children. Students in economically disadvantageous family will actively give up further education if they do not expect enough future return from education. Because of the college enrollment expansion, the college diploma is not as attractive as 10 years ago. And nowadays whether a person can get a job not only depends on the diploma, but also the family resources. Kanbur and Stiglitz (1986) discuss the transition matrices that lead to different level of dynastic inequality.

The results of the article can point directions for inequality-mitigating policies. For example, "National New-Type Urbanisation Plan, 2014–2020" which aims to help left-behind children (those who are left in rural areas while their parents work in urban areas) to enter the urban school system and reunite with their parents could potentially reduce the urban-rural gap, under both direct (better educational resources) and indirect (accompany with parents) influences. The reduction in inequality of opportunity provides evidence in support of the effectiveness of some policies to a certain degree, and we should consider educational inequality an intergenerational problem, and integrated with labor migration, human capital investment, industrial restructuring when making policies.

One of the limitations of using the low-bound estimation is that we underestimate the inequality of opportunity because of the unobserved, omitted, and corelated circumstances/effort. The improvement on accuracy could be realized with increasing amount of future survey data. However, it is unrealistic to include more than 10 circumstance variables because with the sufficient observations under each type, the survey sample is going to be very large, and very difficult to collect.

APPENDIX

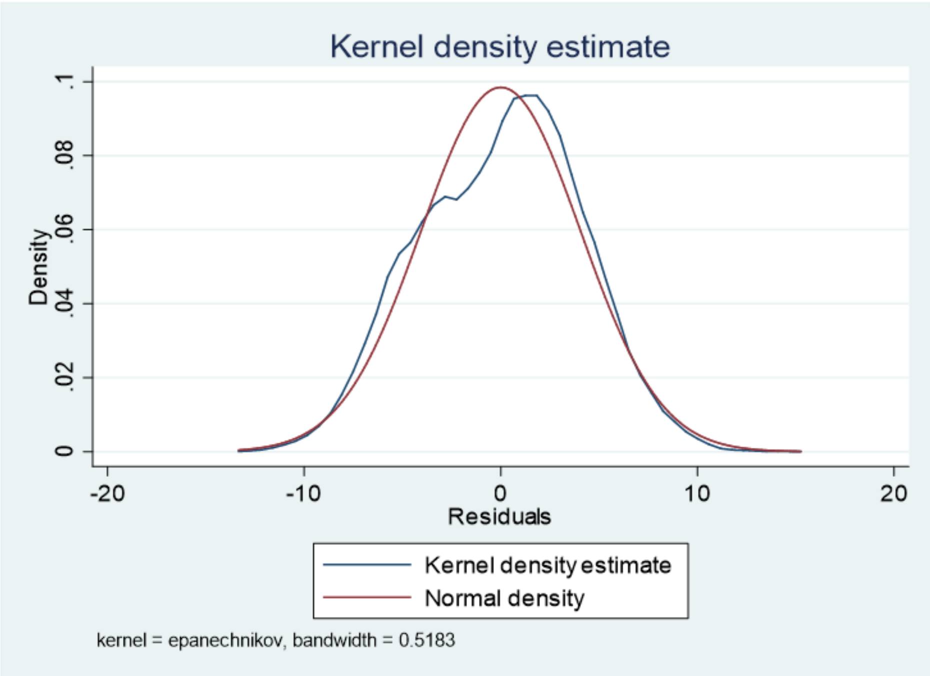


Figure I.A1: kernel density estimate on residual vs. normal density
Source: CFPS2010, 2012, 2014, 2016

```
. sktest res
```

Skewness/Kurtosis tests for Normality

Variable	Obs	Pr(Skewness)	Pr(Kurtosis)	joint adj chi2(2)	joint Prob>chi2
res	17,250	0.0000	0.0000	.	0.0000

Figure I.A2: Normality test for residual
Source: CFPS2010, 2012, 2014, 2016
We can reject the hypothesis that res is normally distributed.

Table I.A1: Multicollinearity test for OLS regression of years of schooling on circumstances

Variable	VIF	1/VIF
Gansu	15.98	0.062567
Henan	15.27	0.065497
Liaoning	13.3	0.075163
Guangdong	10.52	0.095016
Shanghai	9.62	0.103982
Hebei	9.02	0.110842
Shandong	8.13	0.122974
Sichuan	7.27	0.137558
Shanxi	6.93	0.144381

Heilongjiang	5.38	0.18578
Yunnan	5.25	0.190454
Guizhou	4.95	0.202023
Hunan	4.74	0.210955
Shanxi	4.12	0.24253
Three or more siblings (Yes=1)	4.01	0.24953
One or two siblings (Yes=1)	3.87	0.258606
Jilin	3.76	0.266191
Shan'xi	3.75	0.266325
Jiangsu	3.71	0.269663
Guangxi	3.66	0.273535
Anhui	3.44	0.290979
Zhejiang	3.37	0.296705
Hubei	3.05	0.327634
Chongqing	2.27	0.440255
Tianjin	1.97	0.508052
Fujian	1.91	0.522651
Ethnicity (Han=1)	1.34	0.747347
Father-junior high (Yes=1)	1.2	0.835629
Hukou (Urban=1)	1.19	0.842865
Father-primary school (Yes=1)	1.18	0.848248
Father-senior high and over (Yes=1)	1.17	0.851901
Xinjiang	1.05	0.954051
Parents party member (Yes=1)	1.04	0.959141
Hainan	1.03	0.96892
Neimenggu	1.03	0.968963
Qinghai	1.02	0.98322
Gender (Male=1)	1.01	0.993202
Mean VIF	4.64	

Source: CFPS2010, 2012, 2014, 2016

```
. linktest

      Source |       SS           df       MS      Number of obs   =    17,250
-------------+----------------------------------   F(2, 17247)     =   3532.44
       Model |  116010.017         2  58005.0083   Prob > F        =    0.0000
    Residual |  283207.369    17,247  16.4206743   R-squared       =    0.2906
-------------+----------------------------------   Adj R-squared   =    0.2905
       Total |  399217.386    17,249  23.1443786   Root MSE        =    4.0522

------------------------------------------------------------------------------
   yearofsch |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        _hat |   1.050518   .0556224    18.89   0.000     .9414923    1.159543
      _hatsq |  -.0036177   .0038911    -0.93   0.353    -.0112447    .0040092
       _cons |  -.1517548   .1842829    -0.82   0.410     -.512968    .2094583
------------------------------------------------------------------------------

. ovtest

Ramsey RESET test using powers of the fitted values of yearofsch
       Ho:  model has no omitted variables
                F(3, 17209) =      0.77
                   Prob > F =      0.5102
```

Figure I.A3: Model specification test for OLS regression of years of schooling on circumstances.
Source: CFPS2010, 2012, 2014, 2016

Linktest is based on the idea that if a regression is properly specified, one should not be able to find any additional independent variables that are significant except by chance. Linktest creates two new variables, the variable of prediction, _hat, and the variable of squared prediction, _hatsq. The model is then refit using these two variables as predictors. _hat should be significant since it is the predicted value. On the other hand, _hatsq shouldn't, because if our model is specified correctly, the squared predictions should not have much explanatory power. The ovtest performs a regression specification error test (RESET) for omitted variables. It creates new variables based on the predictors and refits the model using those new variables to see if any of them would be significant. Both tests tell us that we do not have a specification error.

```
. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of yearofsch

         chi2(1)      =    21.71
         Prob > chi2  =   0.0000
```

Figure I.A4. Homoscedasticity test for OLS regression of years of schooling on circumstances.
Source: CFPS2010, 2012, 2014, 2016

Large Chi square indicates heteroskedasticity in Breusch-Pagan test.

Determinants of years of schooling											
VARIABLES	all	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89
Hukou (Urban=1)	3.262***	3.968***	3.793***	3.959***	4.221***	3.107***	3.652***	3.338***	3.262***	3.016***	1.767***
Gender (Male=1)	1.492***	2.837***	1.989***	2.446***	2.490***	1.740***	1.410***	1.481***	1.492***	0.136	0.204
Parents party member (Yes=1)	0.0128	0.0267	-0.0722	0.434**	-0.0379	-0.179	-0.294	0.180	0.0128	0.262	-0.225
Ethnicity (Han=1)	0.375***	0.220	-0.545	0.0269	0.233	0.720**	0.344	1.167***	0.375***	1.289***	0.605*
Father-primary school (Yes=1)	2.184***	2.167***	1.755***	1.627***	1.246***	1.195***	1.647***	1.511***	2.184***	1.820***	2.285***
Father-junior high (Yes=1)	3.244***	1.291*	1.533***	0.472	0.932***	1.242***	1.953***	2.475***	3.244***	2.537***	3.095***
Father-senior high and over (Yes=1)	3.878***	0.591	1.416***	1.206***	0.661	2.069***	2.582***	3.330***	3.878***	3.317***	3.962***
One or two siblings (Yes=1)	0.0415	0.618	0.172	-0.117	1.128***	0.632	0.219	0.922**	0.0415	-0.426	-0.824**
Three or more siblings (Yes=1)	-1.016**	0.744*	0.475	-0.152	1.099***	0.548	-0.0868	0.203	-1.016**	-2.134**	-1.959**
Tiajing	-0.370	-9.613**	-2.282	-3.537**	0.890	-1.025	-2.099	0.309	-0.370	-0.403	-0.474
Hebei	-0.974**	-3.803*	-2.370	-0.326	-0.122	-1.919*	-1.848	-1.959*	-0.974**	-1.579	-1.854
Shanxi	-1.223**	-3.965**	-2.407	-1.149	0.258	-2.449**	-1.429	-2.838**	-1.223**	-1.683	-1.740
Neimenggu	-1.497	-	-0.142	-	-0.564	-3.185	-0.616	-	-1.497	-	0.685
Liaoning	-0.708*	-1.979	-1.650	-0.265	-0.233	-2.383**	-1.067	-0.959	-0.708*	-1.927*	-1.879
Jilin	-1.206**	-2.590	-2.149	-1.112	0.0365	-2.811**	-1.724	-2.697**	-1.206**	-1.720	-2.479
Heilongjiang	-1.468**	-4.423**	-3.107**	-1.147	-0.660	-2.311**	-1.903	-2.219*	-1.468**	-2.373**	-1.763
Shanghai	-0.133	-2.918	-1.474	-0.271	1.386	-1.054	-0.267	-0.712	-0.133	0.172	0.474
Zhejiang	-1.408**	-4.169**	-3.940**	-1.658*	-0.821	-2.587**	-1.025	-1.001	-1.408**	-0.459	-1.065
Jiangsu	-0.551	-3.964*	-2.133	-1.337	-0.948	-1.914*	-2.144*	-2.814**	-0.551	1.689	0.342
Anhui	-2.381**	-5.961**	-3.977**	-3.593**	-0.791	-2.902**	-2.632**	-3.427**	-2.381**	-1.196	-2.241
Fujian	-2.891**	-7.918**	-4.900**	-1.448	-2.475*	-5.459**	-4.499**	-4.785**	-2.891**	-3.022**	-2.696
Jinagxi	-2.357**	-4.590**	-3.073*	-2.185**	-1.313	-4.193**	-2.764**	-4.443**	-2.357**	-3.444**	-2.191
Shandong	-1.391**	-3.245	-2.671*	-1.708**	0.354	-2.845**	-1.966	-2.750**	-1.391**	-1.423	-1.700
Henan	-1.177**	-3.605*	-2.566*	-0.758	0.868	-2.434**	-1.651	-2.906**	-1.177**	-1.804*	-2.159
Hubei	-0.863**	-4.774**	-3.648**	-1.494	0.454	-2.069*	-1.291	-1.022	-0.863**	-0.879	-1.999
Hunan	-0.171	-2.283	-1.907	0.182	0.668	-2.054*	0.0671	-0.933	-0.171	0.771	-0.620
Guangdong	-1.367**	-4.671**	-2.832*	-0.885	-0.462	-3.142**	-3.088**	-2.471**	-1.367**	-0.538	-0.244
Guangxi	-1.299**	-3.979*	-4.311**	-0.820	0.683	-2.567**	-1.882	-2.356*	-1.299**	-2.822**	-2.019
Hainan	-1.001	-	-	-	-	-	2.959	-9.856**	-1.001	-	-
Chongqing	-2.086**	-4.319**	-3.199*	-1.731	-1.024	-2.357*	-1.998	-3.516**	-2.086**	-3.640**	-1.619
Sichuan	-2.409**	-5.066**	-3.711**	-2.368**	-2.231**	-3.738**	-1.791	-3.542**	-2.409**	-3.007**	-2.566*
Guizhou	-3.145**	-5.567**	-4.836**	-3.292**	-1.974*	-4.263**	-4.389**	-4.263**	-3.145**	-1.554	-1.046
Yunnan	-2.211**	-6.241**	-4.058**	-1.852*	-1.107	-4.275**	-3.607**	-3.382**	-2.211**	-2.258**	-2.916*
Xizang	-	-	-	-	-	-	-	-	-	-	-
Shan'xi	-0.926**	-2.784	-1.864	-1.093	0.133	-2.529**	-1.783	-2.451**	-0.926**	-1.394	-1.155
Gansu	-2.840**	-5.590**	-3.960**	-2.205**	-1.891*	-3.559**	-3.578**	-4.667**	-2.840**	-3.446**	-2.884**
Qinghai	-	-	-	-	-	-	-	-	-	-	-
Ningxia	-1.189	-	-	-	-	-0.268	-	-	-1.189	-	-
Xinjiang	-5.605	-	-	-	-5.255	-	-	-	-5.605	-	-
Constant	5.307***	4.798**	4.868***	3.128***	1.965*	6.199***	5.711***	5.293***	5.307***	7.575***	8.534***
Observations	19,103	1,052	1,663	2,264	2,106	2,344	2,699	2,459	19,103	1,414	1,407
R-squared	0.325	0.320	0.281	0.325	0.309	0.260	0.297	0.341	0.325	0.424	0.331

Figure I.A5: Determinants of years of schooling.
Source: CFPS2010&2012

Partical contribution to ralative ineqaulity of opportunity

	all	iorp	40-44	iorp	45-49	iorp	50-54	iorp	55-59	iorp	60-64	iorp	65-69	iorp	70-74	iorp	75-79	iorp	80-84	iorp	85-89	iorp
Urban hukou at 12	2	0.202	2	0.220	1	0.142	1	0.1594	1	0.151	1	0.135	1	0.171	3	0.220	2	0.204	1	0.254	3	0.244
Father's education	1	0.161	4	0.259	3	0.222	4	0.275	4	0.278	3	0.201	3	0.192	2	0.213	1	0.161	3	0.269	1	0.158
Province	3	0.250	1	0.211	2	0.190	2	0.235	2	0.222	2	0.178	2	0.189	1	0.205	3	0.252	2	0.268	2	0.208
Parents party member	7	0.322	7	0.319	6	0.277	6	0.325	7	0.309	7	0.259	7	0.296	7	0.341	7	0.325	7	0.424	7	0.330
Male	5	0.301	3	0.246	4	0.225	3	0.248	3	0.241	4	0.214	4	0.272	4	0.315	5	0.301	6	0.423	6	0.329
Number of siblings	4	0.288	5	0.315	5	0.273	7	0.325	6	0.308	6	0.259	5	0.288	5	0.318	4	0.287	4	0.331	4	0.264
Han	6	0.318	6	0.319	7	0.280	5	0.324	5	0.307	5	0.253	6	0.294	6	0.326	6	0.321	5	0.389	5	0.318

Figure I.A6: Partial contribution to relative inequality of opportunity.
Source: CFPS 2010&2012 and author's calculation based on regression results above and method described in section Methods.

REFERENCES

Arneson, R. (1989). Equality and equal opportunity for welfare. Philosophy Studies, 56, 77–93.

Becker, G., & Lewis, H. (1973). On the interaction between the quantity and quality of children. Journal of Political Economy, 81(2), 773–824.

Benos, N., & Zotou, S. (2004). Education and Economic Growth: A Meta-Regression Analysis. World Development, Volume 64, Pages 669-689

Black, S. E., & Devereux, P. J. (2011). Recent developments in intergenerational mobility. Handbook of labor economics, vol. 4, part B (pp. 1487–1541).

Bourguignon, F, "Decomposable Income Inequality Measures," Econometrica, 47(4), 901–20,1979.

Bourguignon, F, Ferreira F, and Menéndez M, "Inequality of Opportunity in Brazil," Review of Income Wealth, 53(4), 585–618, 2007a.

Bourguignon, F, Ferreira F, and Walton M, "Equity, Efficiency and Inequality Traps: A Research Agenda," Journal of Economic Inequality, 5(2), 235–56, 2007b.

Checchi, D. and Vito P. 2010. "Inequality of Opportunity in Italy," Journal of Economic Inequality, 8(4), 429–50, 2010.

Cohen, G. (1989). On the currency of egalitarian justice. Ethics, 99, 906–944.

De Barros, R. P., Ferreira, F.,Molinas, J., & Saavedra, J. (2009). Measuring inequality of opportunity in Latin America and the Caribbean. Washington DC: Palgrave and Mac-Millan and The World Bank.

Delalibera, B., & Ferreira P., (2019) Early childhood education and economic growth. Journal of Economic Dynamics and Control. Volume 98, Pages 82-104.

Dworkin, R. (1981a). What is equality? Part 1: Equality of welfare. Philosophy and Public Affairs, 10, 185–246.

Dworkin, R. (1981b). What is equality? Part 1: Equality of resources. Philosophy and Public Affairs, 10, 283–345.

Emran, M. S., & Sun, Y. (2011). Magical transition? Intergenerational educational and occupational mobility in rural China: 1988–2002 (Online at: http://www.econ.yale.edu/conference/neudc11/papers/paper_055.pdf).

Ferreira, F., Gignoux, J. 2011. The measurement of inequality of opportunity: Theory and an application to Latin America. The review of income and wealth, 2011, Volume 57, Issue 4

Foster, J., Shneyerov, A. "Path Independent Inequality Measures," Journal of Economic Theory, 91(2), 199–222, 2000.

Golley, J., & Kong, S. 2012. Climbing the intergenerational ladder of education in urban, rural and migrant China. Chapter 1) In H., & L. (Eds.), Rebalancing and sustaining growth in China (pp. 2012). Canberra: ANU E Press.

Golley, J., & Kong, S. 2013. Inequality in the intergenerational mobility of education in China. China and the World Economy, 21(2), 15–37.

Golley J., Kong S. 2018. Inequality of opportunity in China's educational outcomes. China Economic Review Volume 51, October 2018, Pages 116-128

Grier K., Hicks D., Yuan, W. 2016. Marriage market matching and conspicuous

consumption in China. Economic Inquiry, 04/2016, Volume 54, Issue 2
Knight, J., Sicular, T., & Yue, X. (2012). Educational inequality in China: The intergenerational dimension. CIBC working paper series, no. 2011–13. University of Western Ontario.
Kunofiwa, T. (2017). Education-economic growth nexus in BRICS countries: Does financial development matter? International Journal of Education Economics and Development. Volume 8, Issue 4, Pages 250-265.
Lefranc, A., Pistolesi, N., & Trannoy, A. (2008). Inequality of opportunities vs. inequality of outcomes: Are western societies alike? The Review of Income and Wealth, 54(4), 513–546.
Li, C. (2010). 'Expansion of higher education and inequality of opportunity in education: A study of the effect of the expansion policy on equalisation of educational attainment' (in Chinese). Sociological Studies, 3, 82–113.
Li, H. J., Zhang, J., & Zhu, Y. (2008). The quality-quantity trade-off of children in a developing country: Identification using Chinese twins. Demography, 45(1), 223–243.
Li, H., Loyalka, P., Rozelle, S.,Wu, B., & Xie, J. (2015). Unequal access to college in China: How far have poor, rural students been left behind? Shanghai finance institute (SFI) working paper series no. SFIWP0046.
Li, Y. (2006). Institutional change and education inequality: Mechanisms in educational stratification in urban China, 1966–2003. Social Sciences in China 497-116 (in Chinese).
Liu, J. (2006). 'Expansion of higher education in China and inequality in entrance opportunities' (in Chinese). Chinese Journal of Sociology, 26(3), 158–179.
Oppedisano, V., Turati, G. 2015. What are the causes of educational inequality and of its evolution over time in Europe? Evidence from PISA. Education economics, 2015, Volume 23, Issue 1
Roemer, John E., Theories of Distributive Justice, Harvard University Press, Cambridge, 1993a.
Roemer, John E., "A Pragmatic Theory of Responsibility for the Egalitarian Planner," Philosophy and Public Affairs, 22(2), 146–66, 1993b.
Roemer, John E., Equality of Opportunity, Harvard University Press, Cambridge, 1998.
Roemer, John E., "Review Essay: The 2006 World Development Report: Equity and Development," Journal of Economic Inequality, 4(2), 233–44, 2006.
Silverstein, M., Cong, Z. 2013. Grandparenting in rural China. Generations, 04/2013, Volume 37, Issue 1
Thompson R. 2014. Stress and Child Development. The Future of Children, 04/2014, Volume 24, Issue 1
Wu, Y., & Huang, C. (2015). Research into urban-rural differences in gender inequality in educational attainment in China, using CGSS 2008 (in Chinese). National School of administration report no. 2 (pp. 41–47).
Xie, Yu and Ping Lu. "The Sampling Design of the China Family Panel Studies (CFPS)." Chinese Journal of Sociology 1(4)(2015): 471-484.
Yu X., Zhang X., et al. 2017. China Family Panel Studies user's manual (3rd edition).
Zhang, Y., & Eriksson, T. (2010). Inequality of opportunity and income inequality in

nine Chinese provinces, 1986–2006. China Economic Review, 21(4), 607–616.
Zhang, X., & Kanbur, R. (2005). Spatial inequality in education and healthcare in China. China Economic Review, 16, 189–204.
Zhou, X., Moen, P., & Tuma, N. B. (1998). Educational stratification in urban China: 1949–94. Sociology of Education, 71(3), 199–222.
Zhou, D; Xu, J; 2017. Heterogeneity in the Intergenerational Transmission of Education and Second Generation Rural-Urban Migrants. International Review of Economics and Finance, November 2017, v. 52, pp. 330-44

CHAPTER II

CONCENTRATION CURVE-BASED ANALYSIS OF CHILDHOOD STUNTING INEQUALITY IN CHINA

- UPDATES FOR THE NEW MILLENNIUM

Introduction

Despite the rapid economic growth in China, inequality in China remains high. The GDP growth rate has remained above 7% annually for the last 3 decades, and income per capital has risen from $990 (1990) to $16,760 (2017)[12]. On the other hand, income inequality has widened, especially between the urban, eastern coastal provinces and rural inland areas due to unbalanced economic development and different access to resources. The World Bank indicates that poverty in China now mainly refers to the rural poor because urban poverty has been largely eliminated due to the economic growth (Appleton et al., 2010). Inequality in education and health are correlated with income inequality, because the poorer population has less access to educational and health resources. The inequality in education and health in turn affects people's human capital and ability to work, and further reinforces the income inequality. Gao et al. (2001), Liu et al. (2002), Zhang and Kanbur (2005) discussed unequal distribution of health service access in China, especially the rural-urban division. Gustafsson and Li (2004), Banister and Zhang (2005) claim the rural poor are "health-stricken" rather than poverty stricken. The Rural Cooperative Medical System (RCMS) has a low annual cost (usually less than $100) but only covers 50% of the treatment fee for regular diseases in 1 year. Populations who suffer from severe or chronic diseases and who could not afford a better medial plan only get limited help from the RCMS. The extreme poor can not afford any medical plan at all. So many poor people are constrained by illness and kept trapped in poverty. Hence, due to its human welfare and economic

---

[12] The World Bank, China Overview. (n.d.). Retrieved from https://www.worldbank.org/en/country/china/overview

development effects, health inequality in China is a timely and important topic to investigate.

Data from the World Health Organization (WHO) show that in 2010, 115 million children were underweight, 55 million children were affected by acute malnutrition, 171 million children under 5 were stunted. From 1990 to 2010, the prevalence of underweight children under 5 in developing countries has fallen from 29% to 18%, but 18% is not sufficient for reaching the goal of reducing underweight children by half (14%) from 1990 to 2015 specified in Millennium Development Goals 1.C [13]. Children's nutritional condition is affected by many environmental factors, many mutually reinforcing, including residence in urban/rural region (hukou in China, Liu et al. 2015; Tian, 2017), family socioeconomic conditions (parents' income, occupation, etc. Wu et al. 2015; Cedraz and Carvalho 1990), sex, age (Zhang et al. 2016) and geographic region of residence (Fu and George, 2015). Smoking and passive smoke also have negative influence on pregnant women and infants. The rate of smokers is higher among low income families. Black et al (2008) estimate childhood malnutrition is the direct reason for 35% of children motility under 5. Each year, over 2 million children under 5 die because of malnutrition globally. Many researchers (Jamison 1986; Zhang et al. 2017; Huang et al. 2013) have found that malnutrition has a negative influence on cognitive development, educational attainment and labor productivity. Childhood malnutrition has a strong association with non-commutable diseases like

[13] MDG 1: Eradicate extreme poverty and hunger. (2015, December 07). Retrieved from https://www.who.int/topics/millennium_development_goals/hunger/en/

hypertension and diabetes in later life. Developmental retardation, iodine deficiency, iron deficiency, and lack of cognitive stimulation become main risk factors for 200 million children who could not exert full potential. Haddad et al (1989) find wage evidence from Philippines that every 1 centimeter increased in height increases 4% of farmer income. Li et al (1999) analyze rural minority children in Yunnan province and claim child malnutrition is caused by long term socioeconomic underdevelopment rather than immediate lack of food.

Analysis of health inequality can reveal how socioeconomic conditions including household income, rural-urban disparity, parents' education, gender and geographic differences affect child malnutrition. According to WHO, measurements of child malnutrition include height-for-age (stunting), weight-for-age (underweight), and weight-for-height (wasting). Most studies focus on wasting and stunting because underweight is less clearly interpreted. In his study, I will focus on stunting because weight can be affected by short term lack of food while stunting score better reflects long-term nutritional conditions. The choice of stunting is also consistent with initial work on health sector inequality with a concentration curve, a modified Lorenz curve to measure health sector inequality. This work by Wagstaff et al (2003) choose stunting as the malnutrition indicator in Vietnam, exploring variables affecting stunting, and decomposing the stunting inequality to show how changes in variable means, variable distribution, and variable's influence on stunting can change the inequality. Chen et al (2007) followed Wagstaff's approach and applying Chinese data through 2000.

The principal objectives of this paper are 1) to measure stunting inequality applying

Wagstaff's approach for more recent data than previous, 2) to investigate the environmental factors that cause child stunting inequality in China, 3) to decompose stunting inequality to find the influential variables, and 4) to discuss policy implications. This article uses Wagstaff et al (2003)'s method and updates the analysis of 1989-2000 data in Chen et al. (2007) from 2000-2011.

Literature

Wagstaff et al. (1989) first introduced the concept of concentration curve, a modification of Gini Coefficient, to measure health inequalities associated with socioeconomic status (SES). This income-related health inequality indicator was designed to measure which health policy instruments better targets low-income population. Wagstaff, van Doorslaer, and Watanabe (2003) further improved the method. They analyzed 1993 and 1998 Vietnamese children's stunting scores (i.e., height-for-age z-scores, HAZ) ranked by per capital household consumption. They found that stunting inequality in both years was largely attributable to inequality of household expenditure and commune level transportation. For each individual in the sample, they decomposed the causes of stunting inequality using linear regression and also decomposed the changes over time by differentiating the resulting regression equation. The noted that the inequality change over time can be caused by changes in determinant variable mean, changes in the importance of the determinant variable to total inequality, and the change in its effects on other variables. They claim that there could be a potential tradeoff between improvement of total inequality and the mean of variable of interest, HAZ scores. Increasing income reduces mean stunting level but increasing inequality in income worsens inequality in health indicator. This same tradeoff is also discussed by Contoyannis and Forster (1999). The tradeoff between decrease of mean stunting level and increase of stunting inequality is worth paying attention to when developing and implementing policies.

Sahn's review (2012) indicates that compared to the absolute level of health, health inequality measurement is multidimensional. He also emphasizes the tradeoff between improving the outcome and the equality of outcome, and he suggests policy makers should take both social justice and economic efficiency into consideration.

Chen, Eastwood and Yen (2007) analyzed data from the China Health and Nutrition Survey (CHNS) for 9 provinces using data from 1989-2000 and chose negative values of the HAZ score as their measurement of malnutrition. Their study concludes that household per capital income, household head's education, hukou in an urban area, and access to bus stop from residence relate to lower level of stunting, while gender and age are not associated with stunting level. They also find unequal distribution of geographic gap and household head's education worsen the inequality in stunting level. This paper builds on their work and investigates what happens to child stunting inequality after 2000 up to 2011.

Methods

A health inequality concentration curve (Figure II.1) is similar to Lorenz curve, expect than it graphs the cumulative percentage of health indicator on the y-axis and the y value is ranked by income on the x-axis from the lowest to the highest. Consider a coordinate axis with cumulative percentage of health received on y-axis and cumulative percent of people ranked by income from the least advantages to the most advantaged on x-axis.

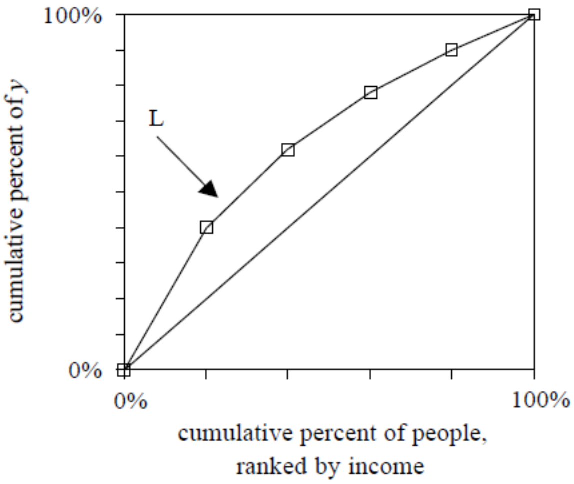


Figure II.1: Concentration curve

If everyone achieves equal health outcomes regardless of their SES, the plot would be a 45-degree line on the coordinate axis. If the curve lies below the 45-degree line, people who are less advantaged in SES suffer more from inequality and vice versa. The further the curve deviates away from the 45-degree line, more unequal the sample is. The Concentration Index, written as C, is twice the area between the concentration curve and diagonal. C can be written in many ways, one of them being:

$$C = \frac{2}{n\bar{y}} \sum_{i=1}^{n} y_i R_i - 1 \qquad (1)$$

where y denotes the health measurement, n denotes the sample size, $\bar{y}$ denotes the sample mean of y, y_i denotes the y for the ith person in the income distribution, and R_i denotes the fractional rank for the ith person. C is the relative measurement of inequality and scaling every individual's health measurement will not change the value of C. C takes the value of 0 if the concentration curve coincides with the diagonal and the sample is completely equal. When the variable is a negative outcome such as stunting measured as HAZ, a disadvantage to the poor implies a concentration curve above the diagonal and negative value of C.

A linear regression model links k determinant x_k to the variable of interest, y:

$$y_i = \alpha + \sum_k \beta_k x_{ki} + \varepsilon_i \tag{2}$$

where α denotes the constant, β denotes the coefficient for the determinant, and ε denotes the error term. In our study, y denotes stunting, and x_k include but not limited to: sex, age, residence, socioeconomic status (family income, expenditure, parents' occupation, education, etc.), mother's age at birth, and etc. By applying the linear model, we assume everyone in the sample faces the same coefficient matrix. Combining equation (1) and (2), the concentration index can be written as[14]:

$$C = \sum_k (\beta_k \bar{x}_k / \bar{y}) / C_k + G_\varepsilon / \bar{y} \tag{3}$$

where $\bar{x}_k$ is the mean of x_k, C_k is the CI for x_k, and G_ε is a generalized concentration index for ε_i, an analogue to the Gini coefficient, defined as:

$$G_\varepsilon = \frac{2}{n} \sum_{i=1}^{n} \varepsilon_i R_i \tag{4}$$

Thus, C can be decomposed to 2 parts: the first part is the weighted sum of the concentration indices of the k determinants x_k, weighted by share for the determinate variable, which is the elasticity of y to the mean of x_k. The second part is the random disturbance term, which defines the factors that influence health inequality but are not captured in the model.

Wagstaff et al. (2003) provides 3 decompositions for change of concentration indices

---

[14] Calculation is proved in Wagstaff, et. al. (2003) "On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam", Appendix A1

over different time period, as:

$$\Delta C = \sum_k (\beta_{kt}\, \bar{x}_{kt}/\bar{y}_t) C_{kt} - \sum_k (\beta_{k(t-1)}\, \bar{x}_{k(t-1)}/\bar{y}_{(t-1)}) C_{k(t-1)} + \Delta(G_{\varepsilon t}/\bar{y}_t) \tag{5}$$

$$\Delta C = \sum_k \eta_k (C_{kt} - C_{k(t-1)}) + \sum_k C_{k(t-1)}\, (\eta_{kt} - \eta_{k(t-1)}) + \Delta(G_{\varepsilon t}/\bar{y}_t) \tag{6}$$

$$\Delta C = \sum_k \eta_{k(t-1)} (C_{kt} - C_{k(t-1)}) + \sum_k C_{kt}\, (\eta_{kt} - \eta_{k(t-1)}) + \Delta(G_{\varepsilon t}/\bar{y}_t) \tag{7}$$

Taking the difference of equation (3) gives equation (5), but this decomposition approach does not tell reader what causes the changes in inequality, whether it's from changes of determinant's mean or changes in its influence on other variables. And empirically, these changes can offset each other. The Blinder–Oaxaca decomposition is a statistical method can decompose the change in means of dependent variable into changes in means of independent variables and change in effect of independent variables (Oaxaca, 1973). Applying a Oaxaca-type decomposition to equation (3) gives equation (6) and (7). These decomposition approaches reveal whether the changes in inequality come from changes in inequality of the determinant variables ($\Delta \mathrm{C}$) or changes in their elasticities($\Delta \eta_k$) .

However, equation (5) ~ (7) are not enough to tell the differences among change of C due to components of η_{kt}. For empirical analysis, it is necessary to know whether the change from η_k comes from β_k or $\bar{x}_k$. Taking total differential of equation (3) gives equation (8)[15], showing changes in $\alpha,\ \beta_k,\ \bar{x}_k, and\ C_k$ accordingly:

$$\mathrm{dC} = \frac{dC}{d\alpha} d\alpha + \sum_k \frac{dC}{d\beta_k} d\beta_k + \sum_k \frac{dC}{d\bar{x}_k} d\bar{x}_k + \sum_k \frac{dC}{dC_k} dC_k + d\frac{GC_\varepsilon}{\mu}$$

$$= -\frac{C}{\mu} d\alpha + \sum_k \frac{\bar{x}_k}{\mu} (C_k - C) d\beta_k + \sum_k \frac{\beta_k}{\mu} (C_k - C) d\bar{x}_k + \sum_k \frac{\beta_k \bar{x}_k}{\mu} dC_k + d\frac{GC_\varepsilon}{\mu} \tag{8}$$

[15] Calculation is proved in Wagstaff, et. al. (2003) "On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam", Appendix A2

From the equation we can see that changing α does impact changes of C, although it does not enter the decomposition. In the case that y represents positive health indicator, μ takes positive value and good health favors population with higher social economic status (C>0), $-\frac{C}{\mu}$<0 ($\frac{dC}{d\alpha}$ <0). Increasing in α indicates an equal increase in every individual's y, and thus reduces relative health inequality condition. An increase in α thus helps the poor to improve their health more, hence reduces the degree of inequality. This is an analog from Podder's article (1993) that equal amount of income increase reduces income inequality. In this study y represents the level of negative HAZ, and thus we would have positive μ and negative C. Increases in α indicate an equal increase in every individual's y (stunting), and $-\frac{C}{\mu}$ is positive. Decreases in α mean equal reduction on every individual's stunting level, and $-\frac{C}{\mu}d\alpha$ <0, so equal reduction on negative health indicator worsens health inequality. This condition is a mirror image of the prior case, and the decrease in α allows the advantageous population to reduce the ill condition more.

Changes of β_k and $\bar{x}_k$ can both affect C directly and indirectly. The direct effect is from β_k and $\bar{x}_k$ themselves, and if C_k=0 (variable is equally distributed), the direct effect is zero. But usually the unequal distribution of the variable changes C through μ indirectly. The indirect effect, could offset the direct effect depending on whether x_k distributes more unequally than y. Let's take an example of positive $\bar{x}_k$ and negative β_k (increase in x_k helps to reduce stunting level), and y represents ill health with a positive μ and negative C. If $\bar{x}_k$ increases with the distribution of x_k unchanged (inequality among x_k unchanged), the direct effect makes the health inequality worse (C becomes more negative). We can observe the change from the equation or follow the logic that inequality inside x_k generates more inequality in y disfavoring the poor. The indirect effect operates through μ. Increases in $\bar{x}_k$ decrease μ, thus making C more negative. In this case, the direct and indirect effects reinforce each other. However, if we consider another case where y represents a positive health outcome with positive μ, C, β_k, $\bar{x}_k$ and C_k (x_k helps to improve good health and distributes favoring the rich). The direct effect of increasing x_k worsens inequality in y (C becomes more positive). On the other hand, increase in $\bar{x}_k$ increases μ, thus reducing the level of inequality in y. in this case, direct and indirect effects could offset each other depending on whether $(C_k - C)$ is positive or negative.

Changes in C_k affects changes in C positively. The term $\sum_k \frac{\beta_k \bar{x}_k}{\mu} dC_k$ indicates positive impact of β_k, $\bar{x}_k$ and negative impact of μ. So in this case, more unequal distributed variable will worsen the HZA inequality.

Data

Data used for this study come from China Health and Nutrition Survey (CHNS)[16], an ongoing project that covers 9 provinces that differ in geographical location, economic development and nature resource endowment in China. This project is implemented by a collaboration of the University of North Carolina at Chapel Hill and the Chinese

[16] This research uses data from China Health and Nutrition Survey (CHNS). We thank the National Institute of Nutrition and Food Safety, China Center for Disease Control and Prevention, Carolina Population Center, the University of North Carolina at Chapel Hill, the NIH (R01-HD30880, DK056350, and R01-HD38700) and the Fogarty International Center, NIH for financial support for the CHNS data collection and analysis files from 1989 to 2006 and both parties plus the China-Japan Friendship Hospital, Ministry of Health for support for CHNS 2009 and future surveys.

Center for Disease Control and Prevention. The survey uses random cluster, multistage, and weighted sampling scheme to draw country samples in each province. More than 4,400 households and 19,000 individuals are covered by the survey, with a high follow up rate. The survey is stratified into individual, family and community level, and I analyze individual survey and household survey data in this study. The survey covers topics including nutrition and physical examination, health service, social economical background, food access, farm production, income and expenditure, family relationship, time allocation at home, media exposure, etc. and each panel is linked by individual ID and household ID. Unfortunately, there are no weights to adjust the data to make them representative for China due to the original survey design[17]. Heilongjiang Province replaced Liaoning Province in 1997, and due to their geographic adjacency, I followed Chen, Eastwood and Yen (2007)'s approach combining these 2 provinces into Northeast. Mega cities including Beijing, Shanghai and Chongqing was added in 2011. For consistency with Chen et al. (2007) these metropolitan areas are not included in the current analysis[18]. CHNS records data starting from 1989 and updates to 2015. Since most panels relevant for this analysis are only updated to 2011, this study will include 9 years of data: 1989, 1991, 1993, 1997, 2000, 2004, 2006, 2009 and 2011. Every observation contains valid entry of positive height value for children under 10-years old (120-month), positive[19] per capita household income (in Yuan, and I take logarithm

---

[17] "Weights for CHNS", Barry M Popkin

[18] See footnote and appendix for 2011 analysis including Beijing, Shanghai and Chongqing variable.

[19] Negative values are eliminated because the regression needs to take logarithm of per capita household income. It turns out that using logarithm of income makes the coefficient more different from 0 than using income, and it also makes the model fits better without changing significant

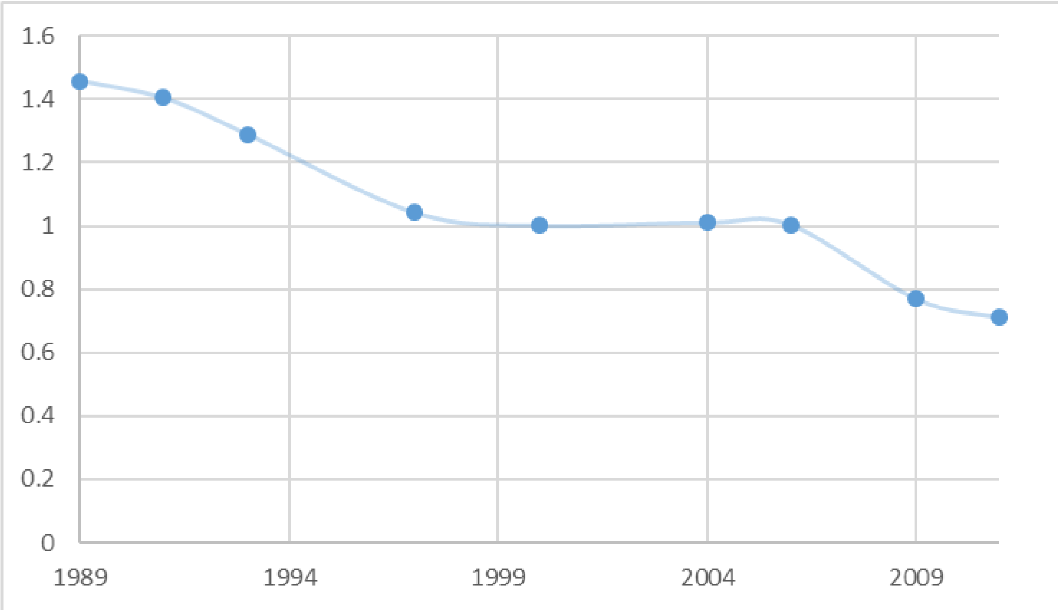
later in regression), household head's education (in years), and dummy variables indicating hukou (urban/rural), gender, and geographic information.

This study follows that approach in Chen et al. (2007) and focuses on children under 10 years old. Wagstaff et al. (2003), Ponce et al. (1998), Martorell et al. (1986), Kostermans (1994) and Chen (2007) all claim that children's height is strongly influenced by genetic factors after 10[20]. Nevertheless, the cutoff age at 5 is more consistent with the majority of literature on stunting (Black et al., 2013; Carlson, 2003; Li, 1999 etc.), and WHO also separate the children into age groups between 0~2, 2~5 and 5~19. The health indicator, stunting (height-for-age) z score is calculated according to World Health Organization (WHO)'s growth reference tables for children from 0 to 2 years, 2 to 5 years and 5 to 19 years. Instead of thinking the z score as "taller the healthier", using negative of z score emphasizes the severity of stunting condition. Hence stunting z score equals to z score times -1. To calculate the cumulative percentage, different signs are eliminated by setting positive z score values to 0. Again, this measurement focuses on the individuals with stunting. For the 9 years of data analyzed, mean stunt z scores have a downward trend (Table II.1 & Figure II.2), which indicates overall better-off of stunting conditions.

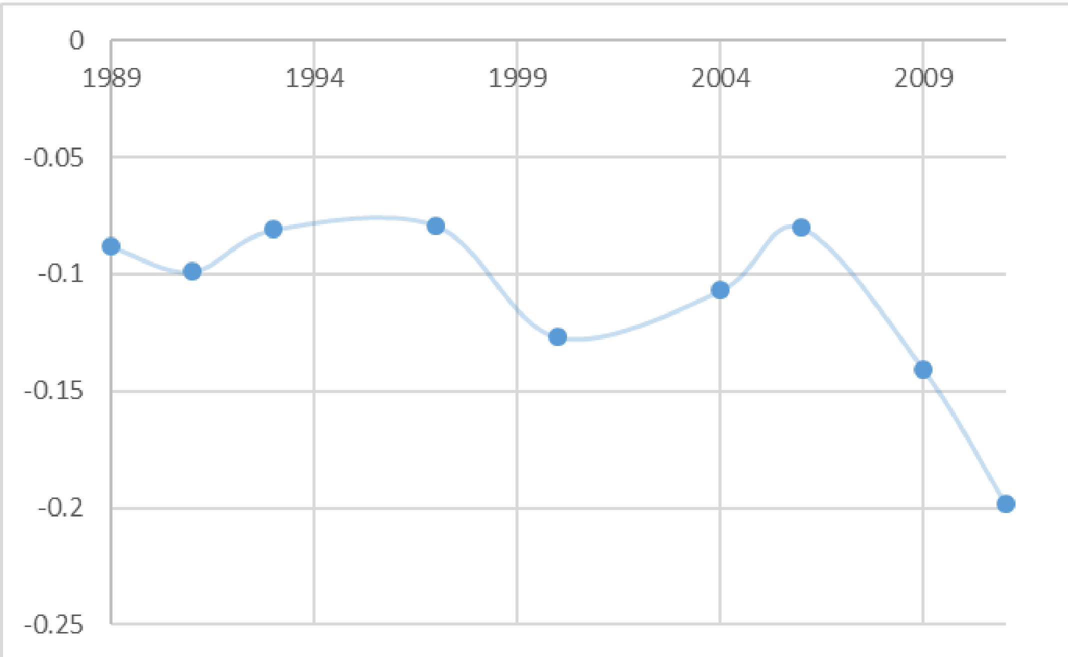
---

level.

[20] Many studies, including WHO, consider height of children under 5 are strongly influenced with nutrition. But this study uses 10 as age cutoff because I want to compare my result with Chen's study, and they used children under 10.



a)



b)

Figure II.2: a) Mean z-score value and b) concentration index, selected years from 1989-2011

Source: CHNS 1989-2011

Table II.1. Sample statistics and concentration indices for available years, 1989-2011[21]

Variables	1989		1991		1993		1997		2000	
	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.
# of obs. (children under 10)	1597		1893		1631		1174		853	
Household income (1000 Yuan)	0.901	0.565	0.876	0.629	1.230	1.188	2.637	2.193	3.242	4.048
Household head's education (year)	6.573	3.943	7.149	3.507	7.250	3.540	6.583	4.028	6.100	4.217
Child's age (month)	39.737	22.763	66.616	31.443	71.589	29.090	77.381	32.872	71.866	33.203
Male	0.535		0.538		0.536		0.549		0.546	
Urban	0.255		0.219		0.210		0.233		0.238	
Jiangsu	0.090		0.088		0.095		0.132		0.110	
Shandong	0.093		0.098		0.099		0.064		0.040	
Henan	0.133		0.114		0.129		0.152		0.127	
Hebei	0.157		0.145		0.151		0.129		0.100	
Hunan	0.109		0.129		0.116		0.086		0.070	
Guangxi	0.149		0.156		0.156		0.133		0.143	
Guizhou	0.137		0.138		0.129		0.145		0.174	
Northeast	0.133		0.132		0.125		0.159		0.237	
Mean Stunting z score	1.458	1.139	1.404	1.034	1.290	0.991	1.043	0.949	1.002	0.997
Concentration index	-0.088		-0.099		-0.081		-0.079		-0.127	

[21] Sample statistics from 1989-2000 is close to Chen, Eastwood and Yen (2007)'s result with minor differences. In the literature they use "community access to bus stop" as an independent variable, but unfortunately community level data is not open to free access.

Table II.1 continue: Sample statistics and concentration indices for 1989-2011 CHNS data

Variables	2004		2006		2009		2011[22]	
	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.
# of obs. (children under 10)	805		734		787		1325	
Household income (1000 Yuan)	4.414	4.615	4.911	7.345	7.714	7.782	12.422	12.815
Household head's education (year)	7.483	3.479	6.831	4.066	7.013	3.812	8.162	4.157
Child's age (month)	66.616	34.159	67.399	34.100	65.919	32.810	60.099	33.731
Male	0.535		0.550		0.560		0.541	
Urban	0.197		0.189		0.211		0.375	
Jiangsu	0.103		0.086		0.107		0.060	
Shandong	0.051		0.084		0.062		0.051	
Henan	0.128		0.128		0.150		0.066	
Hebei	0.071		0.074		0.071		0.048	
Hunan	0.070		0.101		0.129		0.090	
Guangxi	0.164		0.185		0.232		0.159	
Guizhou	0.212		0.192		0.120		0.078	
Northeast	0.200		0.150		0.129		0.070	
Mean Stunting z score	1.012	1.153	1.004	1.179	0.771	1.194	0.712	1.432
Concentration index	-0.107		-0.080		-0.141		-0.198	

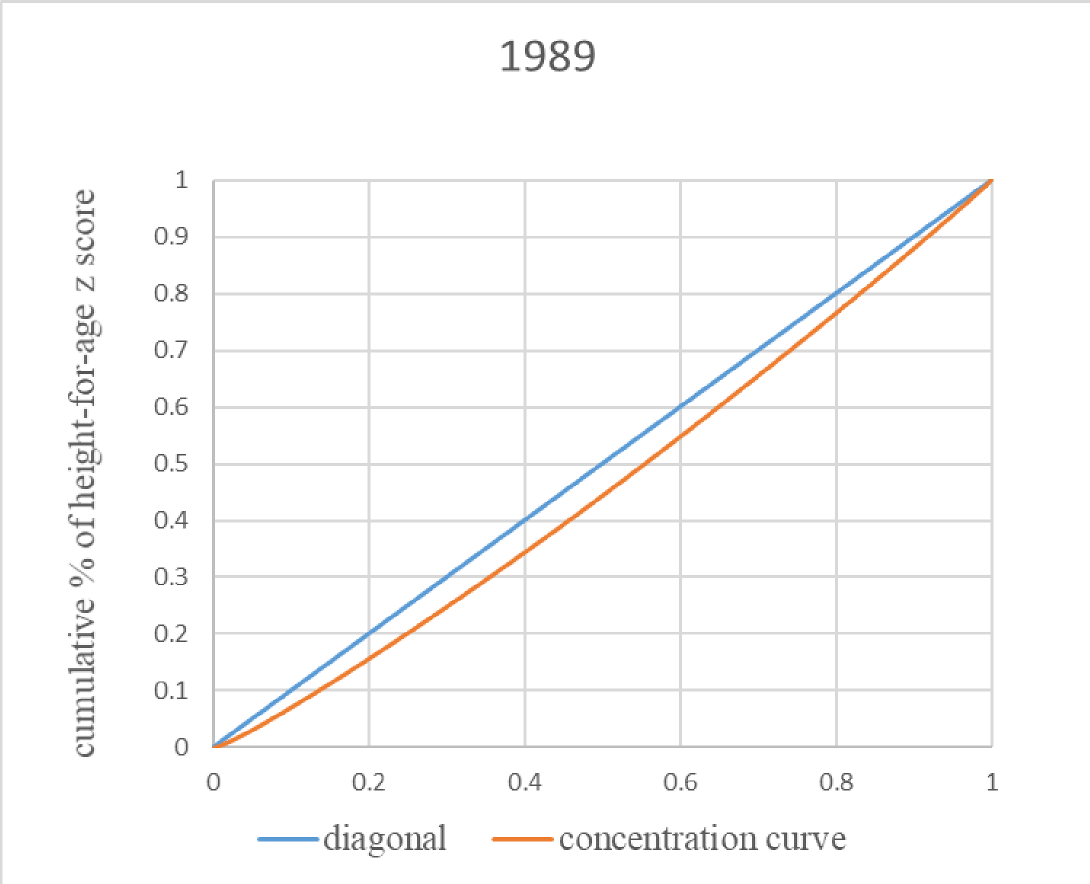
[22] 2011 survey for the first time adds households in mega cities: Beijing, Shanghai, and Chongqing, accounted for 0.106, 0.118, 0.153 of 2011 data that satisfying conditions.

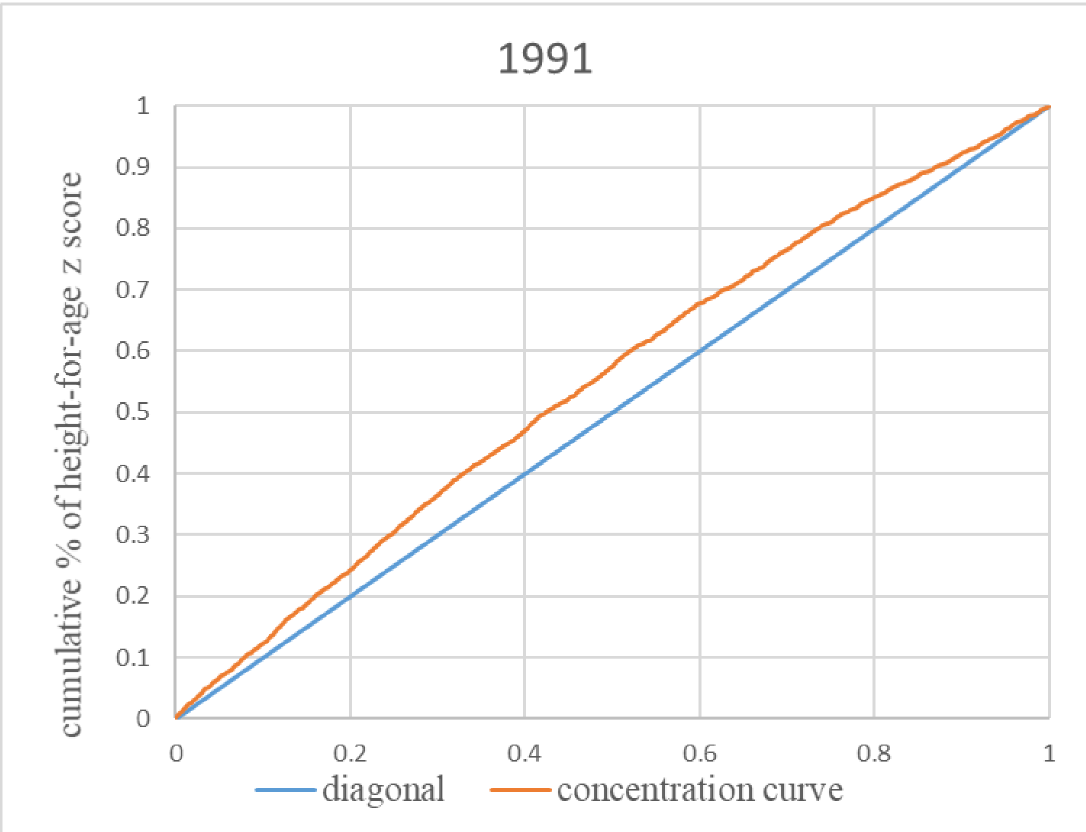
In the CHNS survey, household income is calculated from direct questionnaire response about income, summation of market and non-market activities, including nonmonetary government subsidies, and reposes to questions about expenditures. The data collecting method makes the income data good quality but deviates from the approach used in the China Statistical Yearbook. Chen et al. (2007) find income before adjusting to inflation consist more with national statistics than adjusted income. Another measurement, the Consumer Price Index is also available in adjusted and unadjusted form. However, China's CPI excludes housing price during calculation and considering housing occupies a large portion of household expenditure, this variable is not used to measure household economic status. Wagstaff et al. (2003) used expenditure as the social economic status measure. This study uses income instead of expenditure because of limitation due to expenditure data structure. The response rate for expenditure is low in the data set and including too many 0 entries severally impact the data sensitivity.

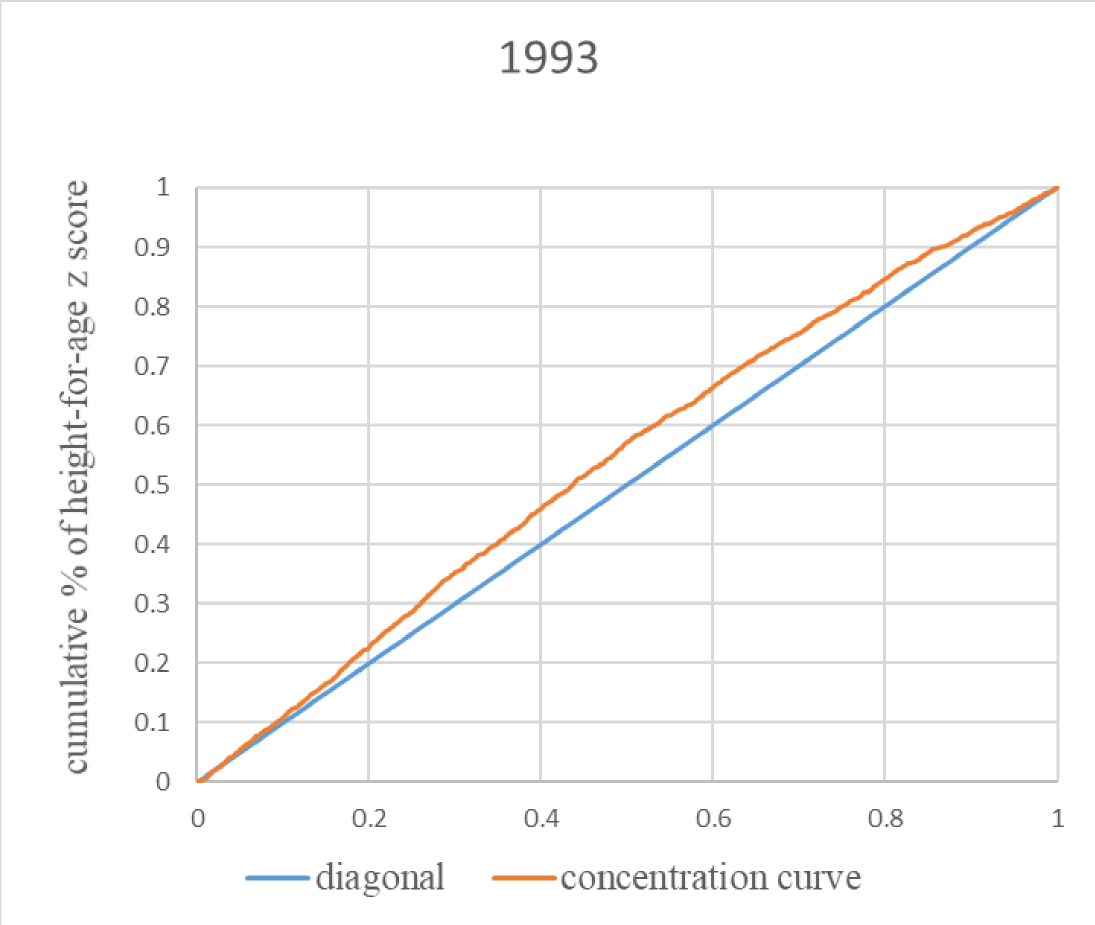
We observe an obvious upward trend in per capita household income (Table II.1), which is consistent with a developing economy and raising price level in the past 20 years in China. The number of observations declines since 1991, and it reflects the reduction of children under 10-years old due to the one-child policy and thus a declining population base. Until 2011, 3 mega cities were covered in the survey and increased the survey response numbers. I use household head's education to represent parents' education level because Chinese family usually have high marriage matching level and education level of parents are highly corelated (Yeoh et al, 2017 and Whyte 1992). Household head's education is relatively constant with standard deviation of about 4 years across all the years. The highest education level in data set is 4^{th} year of college and above, so I use 18 years of education as a cut of for those who receive higher than college level education. Nevertheless, the proportion in this category is small (0.09) among the data set, so using 18 as a cut off has little impact on the results. Children's average age is about 65 months other than 1989, which is only 40 months. Boys' percentage in total sample is constantly a little bit above a half, and urban residents accounts for about 20% throughout all years.

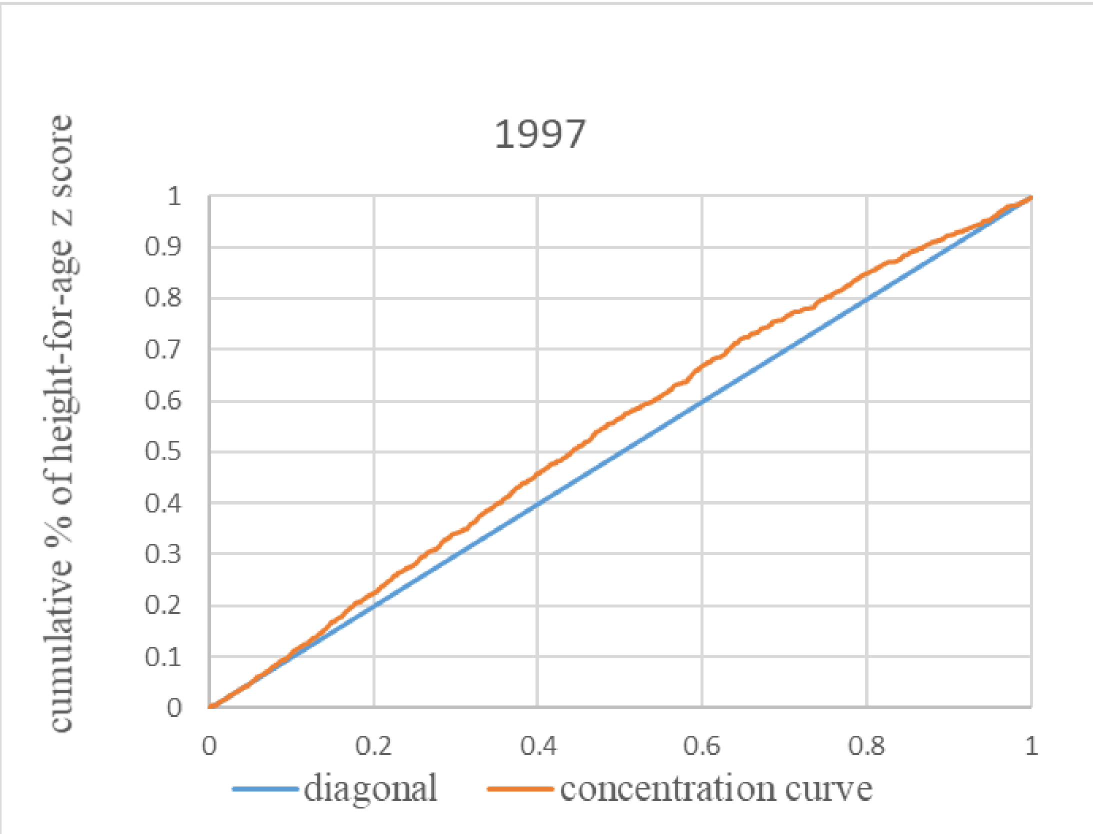
The concentration curves represent stunting z score accumulated by ranking per capita household income from low to high (Figure II.3). The concentration indices are negative throughout all the years, also showing downward trend. Negative indices represent that the poor suffer more from stunting, and larger the absolute value of the concentration index, more unequal among the sample. The concentration indices fluctuate moderately around -0.08 ~ -0.1 from year 1989 to 1997, worsen in 2000, improve slightly in 2004~2006, and deteriorate even more since then (Figure II.2). The variation suggests stunting inequality worsen after 2009. On the other hand, the stunt z-score is declining by time, indicating decrease of total stunting among the sample on average. The lower 10 percentile of concentration curve closely overlay the diagonal, so the inequality among the poorest 10% is subtle, and increase only when income increases (Figure II.3).

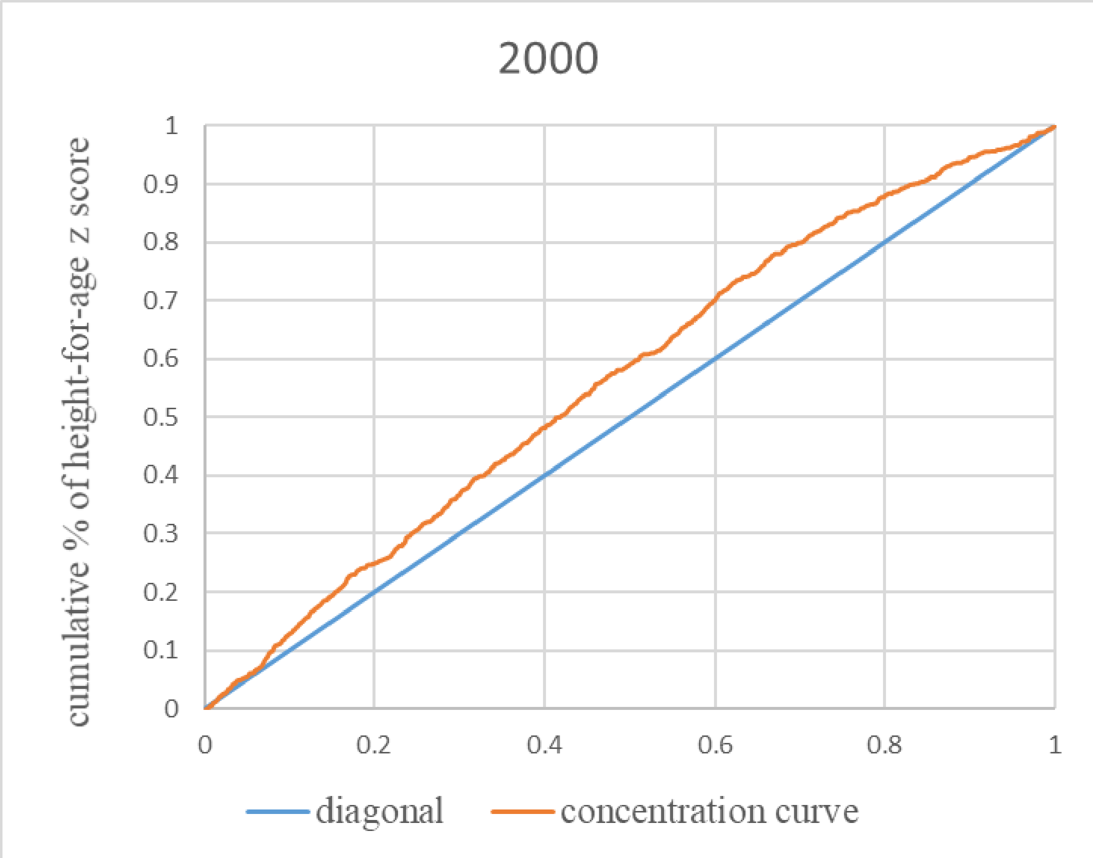
For the specific econometric estimation of equation (2), I assume stunting is a linear function of per capita income (taking logarithm to respond to skewness towards large values), hukou, age (in month), age square, gender, household head's education and geographic location (provincial fixed effect by adding dummies). The explanatory power of the model can be tested by the R-square value. To test the adequacy of OLS model, we need to test for the normality, multicollinearity, model specification, and homoscedasticity. Normality of residuals is only required for valid hypothesis testing and is not required in order to obtain unbiased estimates of the regression coefficients. No multi-collinearity says that there should be no linear relationship between the independent variables. An important implication of this assumption of OLS regression is that there should be sufficient variation in the X's. A model specification error can occur when one or more relevant variables are omitted from the model or one or more irrelevant variables are included in the model. If relevant variables are omitted from the model, the common variance they share with included variables may be wrongly attributed to those variables, and the error term is inflated. On the other hand, if irrelevant variables are included in the model, the common variance they share with included variables may be wrongly attributed to them. Model specification errors can substantially affect the estimate of regression coefficients. OLS assumes that the variance of the error term is constant (homoskedasticity). If the error terms do not have constant variance, they are said to be heteroskedastic. If this variance is not constant, then the linear regression model has heteroscedastic errors and likely to give incorrect estimates. However, heteroskedasticity does not result in biased parameter estimates. Violation of the OLS assumptions will result in overestimated confidence intervals and underestimated p-value, and thus the statistical significance level of the independent variable may also be overestimated.

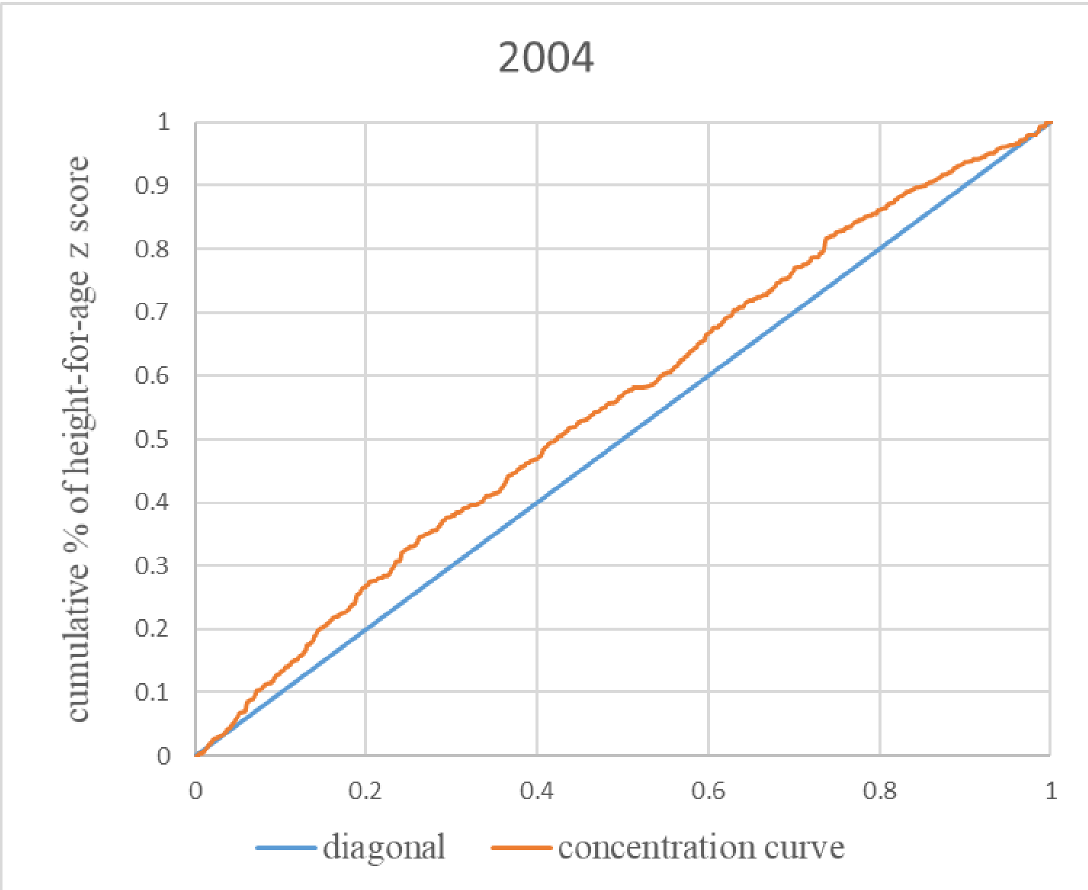

1989
cumulative % of height-for-age z score
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
diagonal
concentration curve

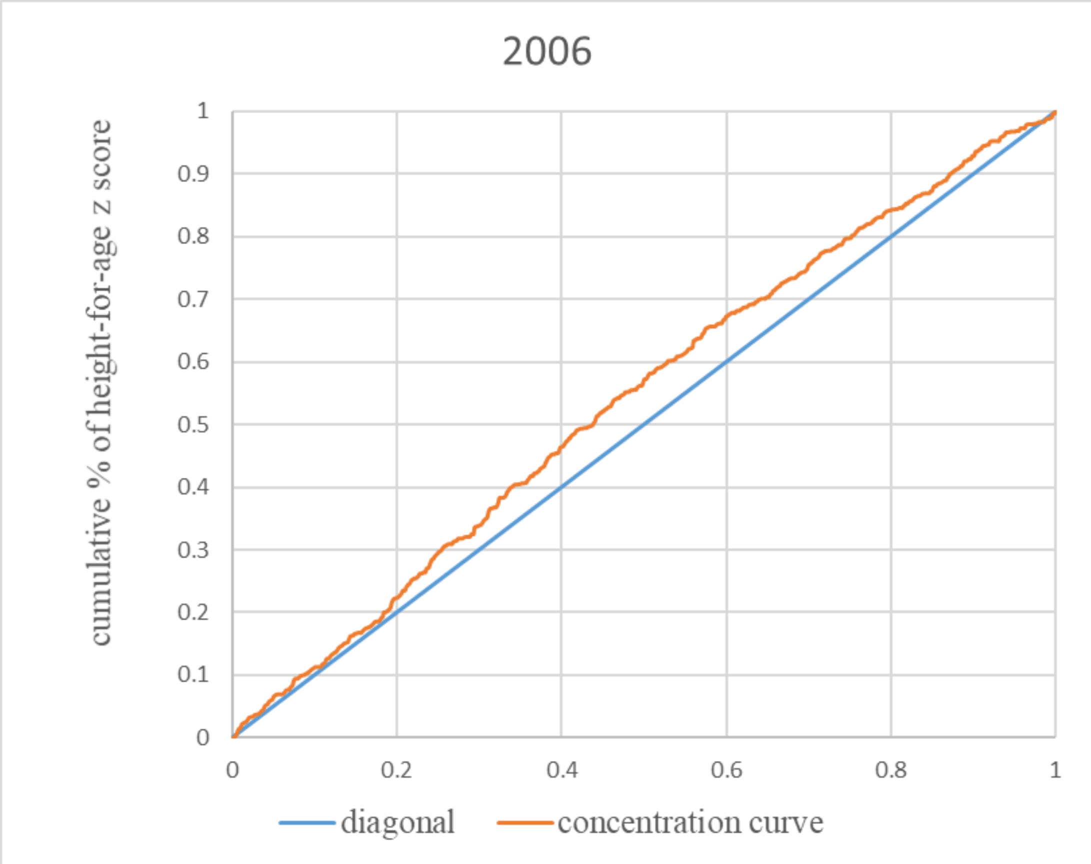

1991
cumulative % of height-for-age z score
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
diagonal
concentration curve

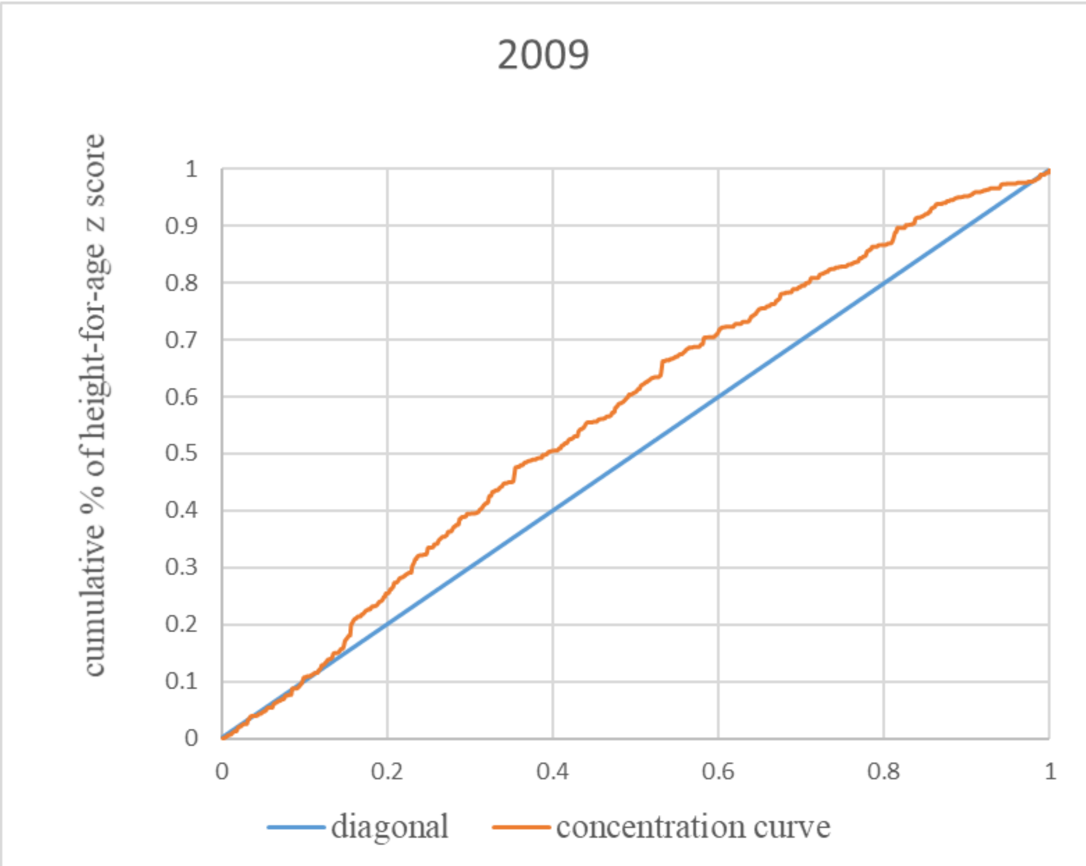

1993
cumulative % of height-for-age z score
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
diagonal
concentration curve


1997
cumulative % of height-for-age z score
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
diagonal
concentration curve


2000
cumulative % of height-for-age z score
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
diagonal
concentration curve


2004
cumulative % of height-for-age z score
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
diagonal
concentration curve


2006
cumulative % of height-for-age z score
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
diagonal
concentration curve


2009
cumulative % of height-for-age z score
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
diagonal
concentration curve

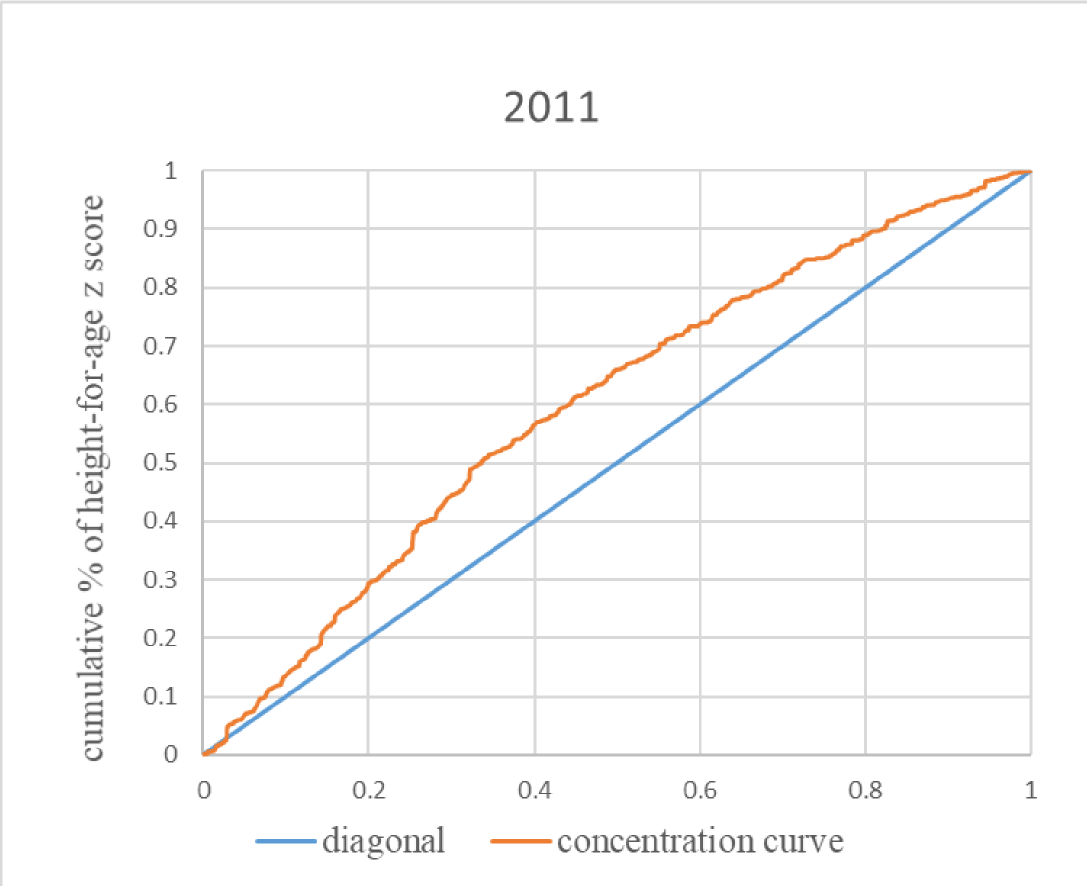


Figure II.3: HAZ ranked by per capita household income (stunting concentration curve) vs. diagonal for available years 1989-2011
Source: CHNS 1989-2011 and author's calculation mentioned in Methods section

Regression results

Table II.2. Similarities and differences between my analysis and Chen el al. (2007)

My analysis	Chen et al. (2007)
Similarities	
The HAZ Concentration Indices are −0.088, −0.099, −0.081, −0.079 and −0.127, respectively, for the five survey years from 1989-2000.	The HAZ Concentration Indices are −0.110, −0.090, −0.067, −0.086 and −0.129, respectively, for the five survey years from 1989-2000.
Per capita household income, urban hukou, and household head's education have negative coefficients and help to reduce stunting throughout 9 survey years. Gender is insignificant explaining stunting.	Per capita household income, urban hukou, and household head's education have negative coefficients and help to reduce stunting throughout 5 survey years. Gender is insignificant explaining stunting.
Compared to Shandong province, residence in Northeast has smaller probability of suffering stunting whereas residence in Hubei, Henan, Hunan, Guangxi Guizhou is the opposite.	Residence in Northeast and Shandong has smaller probability of suffering stunting whereas residence in Guizhou is the opposite.
Per capita household income, urban hukou, household head's education and residence in Guizhou have negative contribution to CI thus disfavor the poor in the 9 survey years. Gender's contribution to CI is 0 across 9 years thus it does not affect stunting inequality.	Per capita household income, urban hukou household head's education, and residence in Guizhou have negative contribution to CI thus disfavor the poor in the 5 survey years. Gender's contribution to CI is 0 across 5 years thus it does not affect stunting inequality.
Differences and updates	
The average stunting HAZ score keeps decreasing in the new decade.	
The HAZ Concentration Indices are −0.107, −0.080, −0.141, and −0.198 respectively for the following survey years from 2004-2010.	
2011-2006 decomposition shows: Log per capita household income disfavors the poor and contributes 49% to ΔC mainly through change in means of log per capita household income; Change in urban hukou worsens the stunting inequality, mainly through change in means of urban hukou.	2000–1993 decomposition shows: The provincial difference disfavors the poor and contributes 45% of ΔC; Log of household income contributes 49% to ΔC mainly through the change in elasticities.
2011-1997 decomposition shows: Log of household income contributes	2000–1991 decomposition shows: The provincial difference disfavors the

53% to ΔC mainly through change in means of log per capita household income; Urban hukou worsens the stunting inequality because the change in means of urban hukou offsets the improvement of its effect on total stunting inequality and its own inequality.	poor and contributes 28% of ΔC; Log of household income contributes 21% to ΔC because of deteriorated income inequality, Urban residence further disfavored the poor because of deteriorated urban-rural division inequality.

Chen et al. found that from 1989 to 2000, (i) provincial fixed effect exists in the model, (ii) coefficients of logarithm of per capita household income have negative signs, and the absolute value of the coefficient declines over year, (iii) coefficients of household head's education have negative signs but are not statistically significant after 1991 (iv) coefficients of children's age have an inverse-U shaped relationship with stunting z score. My findings do not completely agree with Chen's finding because (i) I drop the variable "community access to bus stop" due to data access limitation and (ii) I drop one provincial dummy variable for provincial fixed effect in regression. There are some similarities between the results: other than few exceptions, logarithm per capital income, urban hukou and household head's education years all have negative signs, which indicates these three determinants help to reduce inequality of stunting inequality, which consist with Chen's findings.

The coefficients of log per capita household income on stunting are negative throughout all the years (Table II.3), and 6 out of 9 coefficients are statistically significant. Coefficients for urban hukou are also negative except 2009 (which is not statistically significant), and 7 out of 9 are statistically significant. The absolute values of urban coefficients have a declining trend and indicates shrinking urban-rural difference. This result represents that higher household income and urban Hukou can help reduce the

Table II.3. OLS regression of relationship between HAZ scores and explanatory variables for available years, 1989-2011

Explanatory Variable[23]	1989	t	1991	t	1993	t	1997	t
Log per capita household income	**-0.108**	-3.61	**-0.143**	-5	**-0.114**	-4.22	**-0.066**	-2.04
Hukou (Urban=1)	**-0.454**	-7.25	**-0.452**	-8.2	**-0.344**	-5.78	**-0.253**	-3.99
Children's age in month	**0.045**	10.01	**0.011**	3.55	0.005	1.31	**0.009**	2.47
Children's age square /100	**-0.046**	-8.79	**-0.010**	-4.24	-0.004*	-1.67	**-0.006**	-2.36
Gender (Male=1)	0.0221	0.43	-0.023	-0.53	-0.004	-0.09	-0.002	-0.03
Household's head education in year	**-0.014**	-2.07	**-0.016**	-2.46	-0.010	-1.41	0.003	0.4
Jiangsu	**0.296**	2.45	0.149	1.47	0.005	0.05	0.074	0.59
Henan	**0.262**	2.36	**0.193**	2	0.166	1.66	0.028	0.23
Hubei	**0.456**	4.23	**0.349**	3.87	**0.284**	2.97	**0.352**	2.8
Hunan	**0.645**	5.59	**0.535**	5.8	**0.460**	4.58	**0.403**	2.98
Guangxi	**0.551**	5.08	**0.592**	6.62	**0.378**	3.97	**0.391**	3.14
Guizhou	**0.777**	7	**0.751**	8.07	**0.738**	7.48	**0.857**	6.9
Northeast	-0.149	-1.34	**-0.181**	-1.97	-0.174	-1.76	-0.132	-1.08
Cons	**1.141**	4.9	**2.025**	9.01	**1.882**	8.4	**1.089**	3.88
R^2	0.186		0.174		0.119		0.140	
F for regression	27.86		30.43		16.82		14.51	

[23] Shandong as base Province is omitted.

Table 3 continue

Explanatory Variable	2000	t	2004	t	2006	t	2009	t	2011[24]	t
Log per capita household income	**-0.097**	-2.52	-0.037	-0.89	-0.057	-1.42	-0.054	-1.22	**-0.111**	-2.81
Hukou (Urban=1)	**-0.316**	-4.09	**-0.214**	-2.09	-0.153	-1.36	0.098	0.94	**-0.233**	-2.47
Children's age in month	0.005	1.21	0.005	1.03	-0.002	-0.41	**-0.026**	-5.08	0.007	1.58
Children's age square /100	-0.005	-1.71	-0.006	-1.53	-0.001	-0.27	**0.016**	4.05	**-0.009**	-2.43
Gender (Male=1)	0.011	0.17	-0.076	-1	0.065	0.77	-0.065	-0.79	0.044	0.57
Household's head education in year	-0.007	-0.98	-0.012	-0.99	-0.005	-0.41	**-0.030**	-2.6	-0.011	-1.08
Jiangsu	-0.260	-1.43	-0.053	-0.25	**-0.526**	-2.59	**-0.568**	-2.77	**-0.412**	-2.42
Henan	0.037	0.2	0.325	1.62	0.046	0.24	0.044	0.23	0.329*	1.95
Hubei	0.150	0.8	-0.022	-0.1	-0.035	-0.16	-0.135	-0.61	0.250	1.33
Hunan	0.172	0.88	**0.559**	2.54	0.142	0.73	0.125	0.63	0.191	1.33
Guangxi	0.172	0.97	**0.656**	3.38	**0.355**	2.03	0.006	0.03	0.089	0.73
Guizhou	**0.663**	3.8	**0.862**	4.54	**0.608**	3.51	0.336	1.66	**0.371**	2.37
Northeast	**-0.386**	-2.28	-0.112	-0.59	-0.142	-0.79	-0.424	-2.13	-0.304	-1.86
Cons	**1.777**	4.85	**1.131**	2.84	**1.560**	4.06	2.361	5.39	**1.808**	4.81
R^2	0.186		0.151		0.103		0.113		0.059	
F for regression	14.73		10.86		6.36		7.58		6.30	

Notes: Robust t statistics available; variables with bold text have 5% statistical significance level or better. Due to violation of OLS assumptions in model test, the significance level may be overestimated.

Source: CHNS 1989-2011 and author's calculation $y_i = \alpha + \sum_k \beta_k x_{ki} + \varepsilon_i$

[24] 2011 regression result including Beijing, Shanghai and Chongqing is included in Appendix.

probability of stunting . Child's age has inverted U-shaped relationship with the stunting z-score in 1989~2004, and 2011 but in 2004 the effect is not statistically important. The turning point is around 5 years old, which consists with Chen's and Wagstaff's study. 2006's result is also not statistically important, whereas in 2009 child's age has a significant U-shaped relationship with stunt z-score. Gender's effect to stunting z score are inconclusive because the results have both signs and half of them are statistically insignificant. Inconclusive gender effects may represent limited discrimination between boys and girls, or that the model is not suitable to reflect gender differences. Household head's education (in years) has negative coefficients except 1997, but only 3 of them are statistically significant. The geographic coefficients should be interpreted comparing with the base Province Shandong (omitted in regression). Henan, Hunan, Guangxi and Guizhou have positive coefficients, while Northeast has negative coefficients throughout all years. On the other hand, Jiangsu and Hubei's coefficients have both signs (Hubei's statistically significant coefficients are positive). The result indicates that comparing to Shandong province, children from Northeast have less possibility to suffer from stunting and children from Hubei, Henan, Hunan, Guangxi and Guizhou are more likely to be shorter than the standard height. The result could attribute to different resource endowment and genetic factors among various provinces. Guizhou province is less developed than Shandong province, and it is proven to have statistically significant positive coefficients across all years. Average adult height by provinces from the China Statistic Year Book shows that people from northern China are typically taller than those from southern China , and the fact could explain the negative coefficient for Northeast province. As for Jiangsu, where is geographically adjacent to Shandong, and has very close GDP with Shandong, it's possible that no single factor could outweigh the other, hence the coefficients have both signs. Nevertheless, the provincial coefficient in Henan, Hubei, Hunan, Guangxi and Guizhou are becoming less important after 2000, indicating cross-provincial migration and economic development may surpass the genetic influences.

As mentioned in Methods section, we need to test the OLS model due to various reasons. The R square is relatively low (although not uncommonly low for large cross-sectional datasets), indicating limited ability of this model to explain and predict children's stunning condition. Normality test (Figure II.A1 in appendix) indicates the residual is not normally distributed, thus the standard errors of OLS estimates is not perfectly reliable, which means the confidence intervals would be too narrow. However, normality is not required in order to obtain unbiased estimates of the regression coefficients. Multicollinearity test (Table II.A2 in appendix) proves explanatory variables are not linearly related to each other. As a rule of thumb, a variable whose VIF values are greater than 10 may be problematic. The model specification test (Figure II.A2 in appendix) has a statistically significant p-value and we have to reject the null hypothesis. The result indicates that there exist omitted variables that could have better define the model. The Breusch-Pagan/Cook-Weisberg test for heteroskedasticity (Figure II.A3 in appendix) shows a statistically significant p-value and we have to accept the alternative hypothesis that the variance is not homogenous. Cross-sectional studies are more likely to have heteroscedasticity because they often have values differ in a big range. I already took logarithm of the income but heteroscedasticity still exists. While heteroscedasticity does not cause bias in the coefficient estimates, it does make them less precise. Heteroscedasticity increases the variance of the coefficient estimates and tends to produce p-values that are smaller than the actual value, and in turn a statistically important coefficient could be not statistically important. Thus, the statistical significance of the reported coefficients should be interpreted with caution, because the estimated confidence intervals may be too small in most cases.

Decomposition results

The elasticities and concentration indices of determinants (Tables II.4 and II.5) are calculated based on regression result and by the approach mentioned in Methods section. Gender has no statistically significant influence on stunting inequality (Tables II.6 and II.7).

Table II.4. Elasticity of regression coefficients of each explanatory variable at mean of HAZ scores for available years, 1989-2011

Elasticity	1989	1991	1993	1997	2000	2004	2006	2009	2011
Log per capita household income	-0.479	-0.089	-0.599	-0.478	-0.746	-0.293	-0.450	-0.600	-1.402
Hukou (Urban=1)	-0.080	-0.070	-0.056	-0.057	-0.075	-0.042	-0.029	0.027	-0.123
Children's age in month	1.231	0.513	0.254	0.649	0.358	0.320	-0.150	-2.184	0.614
Children's age square /100	-0.655	-0.368	-0.196	-0.414	-0.323	-0.312	-0.063	1.090	-0.598
Gender (Male=1)	0.008	-0.009	-0.002	-0.001	0.006	-0.040	0.036	-0.047	0.034
Household's head education in year	-0.063	-0.080	-0.054	0.017	-0.045	-0.089	-0.031	-0.269	-0.130
Jiangsu	0.018	0.009	0.000	0.009	-0.029	-0.005	-0.045	-0.079	-0.035
Henan	0.024	0.016	0.017	0.004	0.005	0.041	0.006	0.009	0.030
Hubei	0.049	0.036	0.033	0.044	0.015	-0.002	-0.003	-0.012	0.017
Hunan	0.048	0.049	0.041	0.033	0.012	0.039	0.014	0.021	0.024
Guangxi	0.056	0.066	0.046	0.050	0.025	0.106	0.065	0.002	0.020
Guizhou	0.073	0.074	0.074	0.119	0.115	0.181	0.116	0.052	0.041
Northeast	-0.014	-0.017	-0.017	-0.020	-0.091	-0.022	-0.021	-0.071	-0.030

Note: $\eta_k = \beta_k * \bar{x}_k / \bar{y}$

Source: CHNS 1989-2011 and author's calculation

Table II.5. Concentration Indices of each explanatory variable (ranked by per capita household income) for available years, 1989-2011

Concentration Indices	1989	1991	1993	1997	2000	2004	2006	2009	2011
Log per capita household income	0.079	0.071	0.074	0.062	0.066	0.073	0.077	0.063	0.066
Hukou (Urban=1)	0.341	0.332	0.294	0.299	0.346	0.320	0.267	0.202	0.245
Children's age in month	0.012	0.001	0.017	0.000	0.020	0.008	-0.006	0.027	-0.014
Children's age square /100	0.026	-0.001	0.027	-0.001	0.030	0.018	-0.005	0.042	-0.025
Gender (Male=1)	-0.004	0.002	0.011	-0.004	-0.012	0.011	-0.016	0.011	0.009
Household's head education in year	0.029	0.043	0.036	0.052	0.028	0.088	0.115	0.078	0.109
Jiangsu	0.110	0.141	0.061	0.231	0.441	0.532	0.326	0.351	0.234
Henan	-0.020	-0.201	-0.336	-0.192	-0.376	-0.251	-0.214	-0.202	-0.439
Hubei	-0.132	0.139	-0.074	-0.033	-0.123	-0.028	-0.074	0.130	0.056
Hunan	0.106	0.058	0.096	0.117	0.128	0.125	0.117	0.024	-0.087
Guangxi	0.098	-0.088	0.132	0.024	0.027	-0.113	-0.149	-0.190	-0.312
Guizhou	-0.223	-0.305	-0.061	-0.210	-0.143	-0.200	-0.074	-0.047	-0.159
Northeast	0.004	0.159	0.142	0.092	0.053	0.121	0.166	0.175	0.089

Note: $C_k = \frac{2}{n_k \bar{y}_k} \sum_{i=1}^{n} y_{ik} R_{ik} - 1$, where y_{ik} is the explanatory variable k ranked by per capita household income

Source: CHNS 1989-2011 and author's calculation

Table II.6. Explanatory variables' contribution to Concentration indices for available years, 1989-2011

Explanatory variables' contribution to Concentration indices	1989	1991	1993	1997	2000	2004	2006	2009	2011
Log per capita household income	-0.038	-0.006	-0.044	-0.030	-0.049	-0.021	-0.035	-0.038	-0.093
Hukou (Urban=1)	-0.027	-0.023	-0.016	-0.017	-0.026	-0.013	-0.008	0.005	-0.030
Children's age in month	0.015	0.001	0.004	0.000	0.007	0.003	0.001	-0.059	-0.009
Children's age square /100	-0.017	0.000	-0.005	0.000	-0.010	-0.006	0.000	0.046	0.015
Gender (Male=1)	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	-0.001	0.000
Household's head education in year	-0.002	-0.003	-0.002	0.001	-0.001	-0.008	-0.004	-0.021	-0.014
Jiangsu	0.002	0.001	0.000	0.002	-0.013	-0.003	-0.015	-0.028	-0.008
Henan	0.000	-0.003	-0.006	-0.001	-0.002	-0.010	-0.001	-0.002	-0.013
Hubei	-0.006	0.005	-0.002	-0.001	-0.002	0.000	0.000	-0.002	0.001
Hunan	0.005	0.003	0.004	0.004	0.002	0.005	0.002	0.001	-0.002
Guangxi	0.005	-0.006	0.006	0.001	0.001	-0.012	-0.010	0.000	-0.006
Guizhou	-0.016	-0.023	-0.005	-0.025	-0.016	-0.036	-0.009	-0.002	-0.007
Northeast	0.000	-0.003	-0.002	-0.002	-0.005	-0.003	-0.003	-0.012	-0.003
Residual	-.0126	-.0029	-.4892	-.0137	-.0157	-.0035	-.0010	-.0246	-.0223

Note: Explanatory variables' contribution to Concentration indices = $\eta_k * C_k$

Source: CHNS 1989-2011 and author's calculation

Table II.7. Changes of concentration indices between 2011 and other available years

Changes of concentration indices	11-89	11-91	11-93	11-97	11-00	11-04	11-06	11-09
Log per capita household income	-0.055	-0.086	-0.048	-0.063	-0.043	-0.071	-0.058	-0.055
Hukou (Urban=1)	-0.003	-0.007	-0.014	-0.013	-0.004	-0.017	-0.022	-0.036
Children's age in month	-0.023	-0.009	-0.013	-0.009	-0.016	-0.011	-0.009	0.050
Children's age square /100	0.032	0.015	0.020	0.015	0.025	0.021	0.015	-0.031
Gender (Male=1)	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001
Household's head education in year	-0.012	-0.011	-0.012	-0.015	-0.013	-0.006	-0.011	0.007
Jiangsu	-0.010	-0.009	-0.008	-0.010	0.005	-0.006	0.006	0.020
Henan	-0.013	-0.010	-0.007	-0.012	-0.011	-0.003	-0.012	-0.011
Hubei	0.007	-0.004	0.003	0.002	0.003	0.001	0.001	0.003
Hunan	-0.007	-0.005	-0.006	-0.006	-0.004	-0.007	-0.004	-0.003
Guangxi	-0.012	0.000	-0.012	-0.007	-0.007	0.006	0.003	-0.006
Guizhou	0.010	0.016	-0.002	0.018	0.010	0.030	0.002	-0.004
Northeast	-0.003	0.000	0.000	-0.001	0.002	0.000	0.001	0.010

Source: CHNS 1989-2011 and author's calculation

Inequality in urban hukou (except 2009), per capita household income and household head's education worsen the inequality in child stunting (numerically make concentration indices more negative), thus disfavor the poor. However, the inequality of household head's education has fairly small impact on the inequality of child stunting, accounts for only -0.054 of a total of -0.885 throughout all years. Above results are consistent with Chen et al. (2007). Inequality in age and inequality in age square almost offset each other and have slightly negative contribution to the inequality of child stunting. Residing in Jiangsu favors the poor before 2000 and disfavors the poor afterwards. Residing in Hebei and Guangxi have mixed effects on inequality. Residing in Guizhou, Northeast and Henan except 1989 disfavors the poor, and residing in Hunan expect 2011 favors the poor. The result indicates that children living in poor households in provinces with negative contribution to concentration indices expose to larger possibility of being stunting.

The differences (in Table II.7) between child stunting inequality in 2011 and that in every other year are calculated by subtracting regressors' contribution to concentration indices (from Table II.6). The deterioration of child stunting inequality is mostly caused by change in log of per capita household income and changes in other factors have relatively small effect comparing to household income. The net change of inequalities in child's age is 0.072, and the positive sign indicates the change mitigates the inequality in stunting score. Changes in urban hukou, household head's education (expect in 2009 and 2011), residing in Henan and Hunan have opposite signs and the changes broaden the stunting inequality. Change in gender is negligible.

However, we cannot tell whether the changes in outcome inequality are due to changes in inequality of the determinates or due to changes in elasticities merely based on equation (5) (Table II.6 and II.7). Oaxaca-type decomposition is calculated based on equation (6) and (7). If the absolute value of $\Delta C*\eta$ is larger than $\Delta\eta*C$, it means the change is influenced more by change of inequality rather than change of elasticity, and vice versa. There are 36 combinations of 2 period comparison for 9 years, and I choose 2006 and 2011 because these two years differ the most in concentration index after 2000. The concentration index was improved from 2000 to 2006, and it worsens after 2006. The decomposition analysis for 2006-2011 (Table II.8) indicates that that for residing in Henan, change in elasticity and change in inequality reinforce each other. Other provincial effects are small and in general offset each other. For household income, the improvement in income inequality favors the poor, but could not overweight the changes in elasticities, and changes in income in total contributes to 49% to changes in deterioration of stunting inequality from 2006 to 2011. Changes in urban hukou improve the stunting inequality, mainly comes from changes in elasticities. Change in gender has slightly positive impact on stunting inequality from changes in inequalities of the variable itself. Household head's education contributes to the stunting inequality by changes in elasticities.

Table II.8. Oaxaca-type decompositions for change in inequality 2006 and 2011

Variables	Eq(6)		Eq(7)		Total	
	ΔC*η	Δη*C	ΔC*η	Δη*C	Total	Percent
Log per capita household income	0.015	-0.073	0.005	-0.063	-0.058	49%
Hukou (Urban=1)	0.003	-0.025	0.001	-0.023	-0.022	19%
Children's age in month	-0.005	-0.005	0.001	-0.011	-0.009	8%
Children's age square /100	0.012	0.003	0.001	0.013	0.015	-12%
Gender (Male=1)	0.001	0.000	0.001	0.000	0.001	-1%
Household's head education in year	0.001	-0.011	0.000	-0.011	-0.011	9%
Jiangsu	0.003	0.003	0.004	0.002	0.006	-6%
Henan	-0.007	-0.005	-0.001	-0.011	-0.012	10%
Hubei	0.002	-0.001	0.000	0.001	0.001	-1%
Hunan	-0.005	0.001	-0.003	-0.001	-0.004	3%
Guangxi	-0.003	0.007	-0.011	0.014	0.003	-3%
Guizhou	-0.003	0.006	-0.010	0.012	0.002	-2%
Northeast	0.002	-0.001	0.002	-0.001	0.001	-1%
Residual					-0.030	26%
Total	0.016	-0.103	-0.010	-0.077	-0.117	

Note: $\Delta C = \sum_k \eta_k (C_{kt} - C_{k(t-1)}) + \sum_k C_{k(t-1)} (\eta_{kt} - \eta_{k(t-1)}) + \Delta(G_{\varepsilon t}/\bar{y}_t)$ (6)

$\Delta C = \sum_k \eta_{k(t-1)} (C_{kt} - C_{k(t-1)}) + \sum_k C_{kt} (\eta_{kt} - \eta_{k(t-1)}) + \Delta(G_{\varepsilon t}/\bar{y}_t)$ (7)

Source: CHNS 1989-2011 and author's calculation

I also choose to compare 1997 and 2011 because 1997 has the smallest concentration index while 2011 has the worst stunting inequality (in absolute value). 1997-2011 decomposition result (Table II.9) shows for per capita household income, residing in Jiangsu, Henan, and Northeast, changes in elasticities of the variables and changes in inequalities of variables reinforce each other, and worsen the stunting inequities from 1997 to 2011. The changes in income mainly comes from changes in income elasticity. Inequality of hukou status narrows from 1997 to 2011, but changes in its elasticity offset the improvement and contribute to expansion of stunting inequality. Overall, child's age reduces stunting inequality mainly through changes in age inequality. Gender has no impact on the result. In general, household head's education worsens the stunting inequality by changing in elasticity.

We could explore more details in elasticities which Oaxaca-type decompositions could not tell us: the changes in elasticities might offset each other – the coefficients might decrease while the means of variable might rise. As Wagstaff (2003) mentioned, this decomposition is accurate only for very small changes, hence the results in Table II.7 do not match those in Table II.10 exactly. Log per capita household income and child's age in month differ the most because these two variables account for upper two drivers for inequality change.

Table II.9. Oaxaca-type decompositions for change in inequality 1997 and 2011

Variables	Eq(6)		Eq(7)		Total	
	ΔC*η	Δη*C	ΔC*η	Δη*C	Total	Percent
Log per capita household income	-0.006	-0.057	-0.002	-0.061	-0.063	53%
Hukou (Urban=1)	0.007	-0.020	0.003	-0.016	-0.013	11%
Children's age in month	-0.009	0.000	-0.009	0.000	-0.009	7%
Children's age square /100	0.014	0.000	0.010	0.005	0.015	-12%
Gender (Male=1)	0.000	0.000	0.000	0.000	0.000	0%
Household's head education in year	-0.007	-0.008	0.001	-0.016	-0.015	13%
Jiangsu	0.000	-0.010	0.000	-0.010	-0.010	9%
Henan	-0.007	-0.005	-0.001	-0.011	-0.012	10%
Hubei	0.002	0.001	0.004	-0.002	0.002	-2%
Hunan	-0.005	-0.001	-0.007	0.001	-0.006	5%
Guangxi	-0.007	-0.001	-0.017	0.009	-0.007	6%
Guizhou	0.002	0.016	0.006	0.012	0.018	-16%
Northeast	0.000	-0.001	0.000	-0.001	-0.001	1%
Residual					-0.018	15%
Total	-0.016	-0.085	-0.011	-0.089	-0.119	

Note: $\Delta C = \sum_k \eta_k (C_{kt} - C_{k(t-1)}) + \sum_k C_{k(t-1)} (\eta_{kt} - \eta_{k(t-1)}) + \Delta(G_{\varepsilon t}/\bar{y}_t)$ (6)

$\Delta C = \sum_k \eta_{k(t-1)} (C_{kt} - C_{k(t-1)}) + \sum_k C_{kt} (\eta_{kt} - \eta_{k(t-1)}) + \Delta(G_{\varepsilon t}/\bar{y}_t)$ (7)

Source: CHNS 1989-2011 and author's calculation

Table II.10. Total differential decomposition of changes in inequality

	2006-2011						1997-2011					
	α	β	$\bar{x}$	CI	GC_{ε}	Total	α	β	$\bar{x}$	CI	GC_{ε}	Total
Log per capita household income		-0.042	-0.067	0.003		-0.106		-0.016	-0.087	-0.001		-0.104
Hukou (Urban=1)		-0.005	-0.010	0.001		-0.014		0.002	-0.013	0.003		-0.008
Children's age in month		0.047	0.001	0.001		0.050		-0.009	-0.011	-0.009		-0.029
Children's age square / 100		-0.034	0.001	0.001		-0.031		-0.015	0.011	0.010		0.005
Gender (Male=1)		-0.001	0.000	0.001		0.000		0.002	0.000	0.000		0.002
Household head's education in year		-0.009	-0.001	0.000		-0.010		-0.012	0.001	0.001		-0.010
Jiangsu		0.004	0.006	0.004		0.014		-0.019	-0.002	0.000		-0.021
Henan		-0.005	0.000	-0.001		-0.006		-0.005	0.000	-0.001		-0.006
Hubei		0.000	0.000	0.000		0.000		-0.001	-0.001	0.004		0.002
Hunan		0.001	0.000	-0.003		-0.002		-0.003	0.000	-0.007		-0.010
Guangxi		0.003	0.001	-0.011		-0.007		-0.004	0.001	-0.017		-0.020
Guizhou		0.000	0.000	-0.010		-0.011		0.009	0.007	0.006		0.022
Northeast		-0.006	0.003	0.002		-0.002		-0.004	0.002	0.000		-0.003
Cons	0.02					0.02	0.05					0.05
Residual					-0.030	-0.030					-0.018	-0.018
Total	0.02	-0.046	-0.067	-0.012	-0.030	-0.136	0.05	-0.076	-0.093	-0.010	-0.018	-0.142
Column as %	-15%	34%	50%	9%	22%		-38%	53%	65%	7%	13%	

Note: $\Delta \mathrm{C} = -\frac{C}{\mu}d\alpha + \sum_k \frac{\bar{x}_k}{\mu}(C_k - C)d\beta_k + \sum_k \frac{\beta_k}{\mu}(C_k - C)d\bar{x}_k + \sum_k \frac{\beta_k \bar{x}_k}{\mu}dC_k + d\frac{GC_{\varepsilon}}{\mu}$

Source: CHNS 1989-2011 and author's calculation

For 2006 to 2011 and 1997 to 2011 total differential decomposition, overall, (a) changes in means of variable, (b) changes in their coefficients and (c) changes in their own inequalities all contribute to reinforcing the widening of stunting inequality. In both decompositions, the absolute value of total changes in determinants' inequality is smaller than changes of total stunting inequality. So, the inequality increases in child stunting is larger than inequalities rising in its determinants. We can also assess the magnitudes of these three components contributing to total stunting inequality. For example, in both total decompositions, change in means of log per capita household income contributes more than change in its coefficients and inequalities.

Total differential decompositions (Table II.10) also helps us to explore offsetting effects in elasticities. Sample statistics (in Table II.1) show that sample residing in Northeast province decrease from 15% to 7% from 2006 to 2011. The decrease in $\bar{x}_k$ tends to make C less negative. Meanwhile, the effect of residing in Northeast (β_k) seems to enhance, and in turn worsen the outcome inequality. Oaxaca decomposition (in Table II.8) indicates that the effects roughly cancel out, that $\Delta\eta * C$ is -0.001 in both Oaxaca decompositions using equation (6) and (7). Total differential decomposition (Table II.10) informs us that change in coefficient contributes -0.006 to C (worsen stunting inequality) while decrease in mean rises total inequality by 0.003 and change in variable's own inequality decreases C by 0.002.

Total differential decomposition (Table II.10) could also help to test and prove previous results. For example, we can find that more people get urban hukou, the impact of hukou has declined, and the inequality among population with urban hukou has declines in the decade. From 1997 to 2011, the proportion of urban hukou holder rises from 23% to 38% (in Table II.1), and the rise in mean tends to strengthen the malnutrition inequality. The regression result (in Table II.3, -0.253 in 1997 and -0.233 in 2011) shows the effect of urban hukou on stunting has declined (the absolute values of coefficient get smaller) and the effect seems to narrow the stunting inequality. Concentration indices (in Table II.5) show the urban hukou distributes more equally in 2011 (0.245) than in 1997 (0.299) and the effect should also reduce the inequality level. Total differential decomposition (in Table II.10) shows that change in mean contributes -0.013 towards the total change in C, while change in coefficients and C_k help to offset the total inequality by 0.002 and 0.003 accordingly. The results verify each other.

Conclusion and discussion

The empirical results show that from 1989 to 2011, the average stunting level of China's children under 10 has declined, but the stunting inequality was not reduced with the average improvement of stunting condition. Stunting inequality fluctuates from 1989 to 2006 and worsens in 2009 and 2011. And the inequality among the poorest 10% is negligible.

This article uses different decomposition methods in Wagstaff's article (2003) to explore the causes of inequality and causes of the inequality changes over different time periods. Equation (3) allows us to disclose the inequality of determinants' contribution to the total stunting inequality. Equation (8) allows us to unmask the possible causes for changing inequality over time: the changes could come from (a) changes in means of determinants, (b) changes in the distribution of the determinants, and (c) the changes of their effects on health indicator. The implication of the total decomposition reveals the potential tradeoff between improvement of mean of variable of interest (y) and reduction of stunting inequality. In our case, the notable income growth in China reduces average stunting level. However, increased income exaggerates stunting inequality by increasing income inequality’s contribution to stunting inequality.

From 1989 to 2011, the causal interpretation of regression results indicates household income, urban hukou, household head’s education and residing in Northeast have dampening effect on child stunting level. Among them, the data also shows shrinking urban-rural difference. This result represents that higher household income, urban hukou, higher parents’ education level can help reduce the probability of children malnutrition reflected on their height. The geographic disparity could be attributed to different development level and genetic differences in north and south China (for children older than 60 months).

Decomposition shows inequality in urban hukou (except 2009), per capita household income and household head’s education worsen the inequality in child stunting (make C more negative), thus disfavor the poor. However, the inequality of household head’s education has relatively small impact on the inequality of child stunting.

Total differential decomposition informs us that both in short run (2006 to 2011) and long run (1997 to 2011), the deterioration of stunting inequality is caused by the combined effect from changes in means of variable, changes in their coefficients and changes in their own inequalities. In case of household income, change in means contributes more than change in its coefficients and inequalities.

My analysis confirms and further explains some results from Chen et al. (2007). (i) Income growth does help reduce the average stunting level but worsens inequality mainly by change in its mean. (ii) Household head's education worsens the stunting inequality mostly from the changes in its effect on stunting level. (iii) Gender does not affect significantly in stunting inequality. (iv) Increasing urban hukou contributes positively in stunting inequalities (v) Provincial effects exist and contribute to total inequality.

After 2000, the average stunting score keeps declining during the new decade, whereas stunting inequality ameliorates until 2006 and worsens since then. Log per capita household income, urban hukou, and household's head education continue to contribute to stunting inequality. Residence in Jiangsu is helpful to reduce stunting inequality before 2000 but it starts to attribute to inequality after since.

We observe that household income, urban hukou, household head's education and geographic location are influential factors that cause child stunting inequality. Compared to more developed countries, policies in China aims to mitigate children nutritional inequality is in the beginning phase. In US, school lunch program, Food Stamp program and Special Nutrition Supplement program for Women, Infants and Children (WIC) have national coverage and have helped to improve nutrition status among the low-income population (Carlson, 2003). China has no specific nutrition program towards children and pregnant women before 2011. Starting in 2011, State Council of China announced "Rural China compulsory education students nutrition improvement program", and provided meal subsidies for 26 million students at school in 680 selected countries (cities) with ¥3 ($0.45) per person per day (250 days per year, ¥4 for primary students and ¥5 for middle school students with financial difficulties). Until the end of 2013, the government has spent over ¥30 billion on the program and provided for 32 million rural students. Also in 2011, charity organizations like China Social Welfare Foundation worked with domestic well-known social media, journalists and philanthropists initiated Free Lunch for Children mutual fund program and proposed for donation of ¥3 for each children (raised to ¥4 in 2016). Until 2017, the donation has passed ¥333 million and helped 893 schools. However, both programs have limitations. First, the amount of each meal subsidy is way too low considering current food price. Second, the coverage of the program is limited by geographic conditions. Third, these two programs only provide meal subsidy to school students but do not cover infants. China has welfare programs[25] targeting the poor but there is no specific program providing necessary nutrition for the pregnant women and infants. To test if the programs effective, future studies can work on data after 2011 if released. Also, residual part constitutes approximately 20% in the short run change (2006~2011) and 10% in the long run change (1997~2011), which indicates additional variables could be explored on affecting children stunting.

Other than that, the health indicator itself reflects inequality limited to surviving children. It does not count for child mortality, especially if gender specific. Also, height

[25] For example, Wubaohu program covers population (families) with elderly, disabled and orphaned minors without income source, and provides free food, clothes, fuel, housing, and health care. (Ministry of Civil Affairs, 1994)

only presents general absorption of nutrition, but does not fully reflect health conditions related to diseases, especially current ones. For example, in Ta-liang Mountains area inside Szechwan province, children suffer from high rate of HIV infection due to its location (Ta-liang Mountains area is one of the least economic developed areas, extremely poor in resources, has very limited transportation access, and it is China's biggest drug-dealing zone).

APPENDIX

Table II.A1. 2011 data including mega cities

Variable	β	η	C_k	Regressors' contribution to C_k
Log per capita household income	-0.064	-0.806	0.066	-0.093
Hukou (Urban=1)	-0.184^{*}	-0.097	0.245	-0.030
Children's age in month	0.007	0.591	-0.014	-0.009
Children's age square /100	-0.009^{**}	-0.600	-0.025	0.015
Gender (Male=1)	0.046	0.035	0.009	0.000
Household's head education in year	-0.001	-0.011	0.109	-0.014
Jiangsu	-0.470^{**}	-0.040	0.234	-0.008
Henan	0.329	0.030	-0.439	-0.013
Hubei	0.216	0.015	0.056	0.001
Hunan	0.159	0.020	-0.087	-0.002
Guangxi	0.087	0.019	-0.312	-0.006
Guizhou	0.372^{*}	0.041	-0.159	-0.007
Northeast	-0.355	-0.035	0.089	-0.003
Shanghai	-0.377^{*}	-0.056	0.548	-0.031
Beijing	-0.293	-0.049	-0.159	0.008
Chongqing	0.255	0.055	0.449	0.025
Cons	1.329^{***}	-		
R^2	0.071	-		
F for regression	6.26	-		
Total				-0.166

```
. swilk resid

                   Shapiro-Wilk W test for normal data

    Variable |        Obs       W           V         z       Prob>z
-------------+------------------------------------------------------
       resid |        853    0.94271     31.265     8.469    0.00000
```

Figure II.A1. Normality test for OLS regression of HAZ scores on explanatory variables, 2000.

Table II.A2. Multicollinearity test for OLS regression of HAZ scores on explanatory variables, 2000.

Variable	VIF	1/VIF
agesq	19.54	0.05117
ageinmonth	19.49	0.0513
Northeast	5.4	0.185333
Guizhou	4.54	0.220122
Guangxi	3.98	0.251044
Henan	3.85	0.25971
Jiangsu	3.39	0.295191
Hubei	3.26	0.307102
Hunan	2.6	0.384139
log_hhinc_pc	1.28	0.783662
urban	1.12	0.889554
heduyear	1.04	0.96362
genderM1F0	1.02	0.978399
Mean VIF	5.42	

```
      Source |       SS           df       MS      Number of obs   =       853
-------------+----------------------------------   F(2, 850)       =     99.14
       Model |  160.110461         2  80.0552305   Prob > F        =    0.0000
    Residual |  686.359751       850   .80748206   R-squared       =    0.1892
-------------+----------------------------------   Adj R-squared   =    0.1872
       Total |  846.470212       852  .993509638   Root MSE        =     .8986

------------------------------------------------------------------------------
        adjz |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        _hat |   .4603006   .2966105     1.55   0.121    -.1218742    1.042475
      _hatsq |   .2600056   .1386628     1.88   0.061     -.012156    .5321673
       _cons |   .2317871   .1462131     1.59   0.113    -.0551939    .5187682
------------------------------------------------------------------------------


. ovtest

Ramsey RESET test using powers of the fitted values of adjz
       Ho:  model has no omitted variables
                 F(3, 836) =      2.84
                  Prob > F =      0.0368
```

Figure II.A2. Model specification test for OLS regression of HAZ scores on explanatory variables, 2000.

If the model is specified correctly, in linktest, _hat should be significant and _hatsq shouldn't. The ovtest suggests we need to reject the null hypothesis and the model has omitted variables.

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of adjz

         chi2(1)      =    25.07
         Prob > chi2  =   0.0000
```

Figure II.A3. Homoscedasticity test for OLS regression of HAZ scores on explanatory variables, 2000.

Large Chi square indicates heteroskedasticity in Breusch-Pagan test.

REFERENCES

Appleton, S., Song, L., Xia, Q., (2010). Growing out of Poverty: Trends and Patterns of Urban Poverty in China 1988–2002, World Development. Volume 38, Issue 5, Pages 665-678

Banister, J., & Zhang, X. (2005). China, economic development and mortality decline. World Development, 33(1), 21−41.

Black RE, Victora CG, Walker SP, et al. and the Maternal and Child Nutrition Study Group . Maternal and child undernutrition and overweight in low-income and middle-income countries. Lancet. 2013;382:427–451.

Carlson, A., & Senauer, B. (2003). The impact of the special supplemental nutrition program for women, infants], and children on child health. American Journal of Agricultural Economics, 85(2), 479−491.

Chen, Z., Eastwood, D., Yen, S. 2007. A decade's story of childhood malnutrition inequality in China: Where you live does matter. China Economic Review 18, 139-154.

Contoyannis, P., Forster, M., 1999a. Our healthier nation? Health Economics 8, 289–296.

Contoyannis, P., Forster, P., 1999b. The distribution of health: a theoretical framework. Journal of Health Economics 18 (5), 605–622.

Cedraz, L M; Carvalho, F M., 1990. Family income and child malnutrition in the coast of Camaçari, Brazil. Archivos latinoamericanos de nutricion, Volume 40, Issue 3

Deaton, A., 1997. The Analysis of Household Surveys: A Microeconometric Approach to Development Policy. Johns Hopkins University Press, Baltimore, MD.

Ettner, S. L. (1996). New evidence on the relationship between income and health. Journal of Health Economics, 15(1), 67−85.

Fu, Q., George, L., 2015. Sex, Socioeconomic and Regional Disparities in Age Trajectories of Childhood BMI, Underweight and Overweight in China. Asian Population Studies. Jul2015, Vol. 11 Issue 2, p134-148. 15p.

Gao, J., Tang, S. L., Tolhurst, R., & Rao, K. Q. (2001). Changing access to health services in urban China: implications for equity. Health Policy and Planning, 16(3), 302−312.

Haddad, Lawrence J., and Howarth E. Bouis. "The Impact Of Nutritional Status On Agricultural Productivity: Wage Evidence From The Philippines*." Oxford Bulletin of Economics and Statistics, vol. 53, no. 1, 2009, pp. 45–68., doi:10.1111/j.1468-0084.1991.mp53001004.x.

Huang, C., et al. 2013. Malnutrition in early life and adult mental health: Evidence from a natural experiment. Social Science & Medicine Volume 97, Pages 259-266

Jamison, D. T. (1986). Child malnutrition and school performance in China. Journal of Development Economics, 20(2), 229−309.

Kakwani, N.,Wagstaff, A., & van Doorslaeff, E. (1997). Socioeconomic inequalities in health: measurement, computation, and statistical inference. Journal of Econometrics, 77(1), 87−103.

Kostermans, K., 1994. Assessingthe quality of anthropometric data. LSMS WorkingPaper No. 101. The World Bank, Washington, DC.

Liu, H., Rizzo, J., Fang, J., 2015, Urban-rural disparities in child nutrition-related health outcomes in China: The role of hukou policy, BMC Public Health. 15.10763
Liu, G. G., Zhao, Z., Cai, R., Yamada, T., & Yamada, T. (2002). Equity in health care access to: assessing the urban health insurance reform in China. Social Science & Medicine, 55(10), 1779−1794.
Li, Y., Guo, G., Shi, A., Li, Y., Anme, T., & Ushijima, H. (1999). Prevalence and correlates of malnutrition among children in rural minority areas of China. Pediatrics International, 41(5), 549−556.
Liu, Y., Hsiao, W. C., & Eggleston, K. (1999). Equity in health and health care: the Chinese experience. Social Science & Medicine, 49(10), 1349−1356.
Ministry of Civil Affairs, the People's Republic of China., (1994). Directive 141: Provisions for Wubaohu in rural area.
Oaxaca, R.L., 1973. Male–female wage di&erentials in urban labor markets. International Economic Review 14, 693–709.
Podder, N., 1993. The disaggregation of the Gini coeKcient by factor components and its applications to Australia. Review of Income and Wealth 39 (1), 51–61.
Ponce, N., Gertler, P., Glewwe, P., 1998. Will Vietnam grow out of malnutrition? In: Dollar, D., Glewwe, P., Litvack, J. (Eds.), Household Welfare and Vietnam's Transition. World Bank, Washington, DC, pp. 257–275.
Rao, V.M., 1969. Two decompositions of concentration ratio. Journal of Royal Statistical Society Series A 132 (3), 418–425.
Sahn, D.E., Younger, S.D., 2000. Expenditure incidence in Africa: microeconomic evidence. Fiscal Studies 21 (3), 329–348.
Tian, T., 2017. Urban-Rural Disparities in Child Overweight and Obesity in China: The Role of Hukou Policy. Georgetown University, ProQuest Dissertations Publishing
The University of North Carolina, Carolina Population Center (UNC-CPC) (2019). Constructed variables,
http://www.cpc.unc.edu/projects/china/data/datasets/convar.html
van Doorslaer, E., & Jones, A. M. (2003). Inequalities in self-reported health: validation of a new approach to measurement. Journal of Health Economics, 22(1), 61−87.
Wagstaff, A. (2005). Decomposing changes in income inequality into vertical and horizontal redistribution and reranking, with applications to China and Vietnam. World bank policy research working paper, Vol. 3559. Washington, DC: The World Bank
Wagstaff, A., & Watanabe, N. (2003). What difference does the choice of SES make in health inequality measurement? Health Economics, 12(10), 885−890.
Wagstaff, A., van Doorslaer, E., & Paci, P. (1989). Equity in the finance and delivery of health care: some tentative crosscountry comparisons. Oxford Review of Economic Policy, 5(1), 89−112.
Wagstaff, A., van Doorslaer, E., & Watanabe, N. (2003). On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam. Journal of Econometrics, 112(1), 207−223.
Whyte, Martin King. From Arranged Marriages to Love Matches in Urban China. Hong Kong Institute of Asia-Pacific Studies, Chinese University of Hong Kong, 1992.
http://www.mca.gov.cn/artical/content/WJYL_SHFL/200481990113.html, Accessed

April 2019.
Wu, L., Yang, Z., Yin, S., Zhu, M., Gao, H., 2015. The relationship between socioeconomic development and malnutrition in children younger than 5 years in China during the period 1990 to 2010. Asia Pacific journal of clinical nutrition. Volume: 24 Issue: 4 Page: 665
Yeoh, B., Chee, H., Baey, G., 2017, Managing risk, making a match: brokers and the management of mobility in international marriage. Mobilities. Apr2017, Vol. 12 Issue 2, p227-242. 16p.
Zhang, N., Becares, L., Chandola, T., 2016. Patterns and Determinants of Double-Burden of Malnutrition among Rural Children: Evidence from China. Published: July 8, 2016 https://doi.org/10.1371/journal.pone.0158119
Zhang, X., & Kanbur, R. (2005). Spatial inequality in education and health care in China. China Economic Review, 16(2), 189−204.
Zhang, Z., Liu, J., Li, L., 2017. The Long Arm of Childhood in China: Early-Life Conditions and Cognitive Function Among Middle-Aged and Older Adults. Journal of Aging and Health. Volume: 30 issue: 8, page(s): 1319-1344

CHAPTER III

ASSESSING INTRA-HOUSEHOLD INEQUALITY IN BODY MASS INDEX FOR CHINA DURING 1989-2011

Introduction

When data are collected at the household level rather than individual level, intra-household inequality can be ignored, and the total inequality in health and nutrition outcomes can be underestimated. However, specific assessment of intra-household inequality, especially for health outcomes, is an underdeveloped research area (Kanbur, 2016). To the best of my knowledge, this study is the first to focus specifically on intra-household health inequality in China. I am especially interested in whether women and children are thinner than male adults in the family, and whether the outcomes change with household income levels. The objectives of this article are 1) to assess the inequality of intra-household well-being as represented by body mass index (BMI), 2) to justify the choice of BMI as well-being indicator, 3) to explore whether the health resource depravation exists among family members, and 4) to test the hypothesis of a Kuznets curve relationship between intra-household inequality and measurement of well-being.

This article measures health inequality with the BMI, which is calculated by one's weight in kilograms divided by height squared in centimeters. BMI in a certain range suggest better health condition and BMI outside the thresholds can indicate health problems. Individuals with low BMI may indicate insufficient nutritional intake and/or in other health conditions. On the other hand, BMI higher than the upper limitation points to health problems linked to obesity. Thus, BMI reflects individual access and consumptions over health-related resources, including food, health care and sanitation. Advantages and disadvantages of using BMI as health indicator are discussed in the

Methods section. In this article I treat BMI analogous to income in the standard inequality literature: well-being is measured by BMI and inequality is measured by its dispersion in distribution. Using individual level and household level well-being indicators bias the measurement of inequality. Therefore, I am interested in the intra-household inequality calculated based on a non-income well-being indicator, and the extent to which it is attributable to the total inequality.

Intra-household inequality is important because it affects policies in various ways. Underestimation of total inequality by omitting intra-household inequality distorts the achievement of inequality/poverty reduction programs. Resource allocation patterns inside the family influences the effect of minimum wage polices and public transfer programs because the increase in resources does not necessarily go to targeted individuals due to family reallocation. One of the underlying reasons is that intra-household resource allocation can be a function of technology (ability to turn marginal calories into income) and preferences (Haddad et al., 1995). I will explore if families will preferentially allocate health-related resources to relatively more vulnerable family members or away from them, as family resource constraints decrease through higher income. The effect will be measured in the form of BMI share among household members.

In this paper I will also explore whether Kuznets curve exists between intra-household inequality and measurement of well-being (household average BMI). Kuznets's seminal work (1955) proposes an inverted-U relationship between income inequality and income development. The hypothesis claims that countries in the beginning stage of economic growth have little inequality, and the inequality widens as the economic development, and begins to decline when reaching a certain level of growth. The hypothesis is tested by many researchers using different countries' data, and the results show both supportive and contradictory evidence.

Literature Review

Decision-making inside families determines resource distribution among family members. Many in the literature including Browning, Chiappori and Weiss (2014) discuss the complex decision-making processes and their internal mechanism. The dominate analytical approach, the unitary model of household maximizes the family utility function under single budget constraint. Other models include cooperative and non-cooperative models. Various distribution approaches induce intra-household well-being inequalities in health outcome; BMI in our case. Kanbur (2016) explains that the unitary model causes intra-household inequality because the family utility function is not necessarily inequality averse, while inequality rises in Nash bargaining game under cooperative model. Empirically, the unitary model that implies equal shares contrasts with the implications of cooperative and non-cooperative bargaining models (Browning and Chiappori, 1998; McElroy, 1990). Inevitably, unequal allocation of family resources causes intra-household inequality.

Many previous works regarding intra-household inequality find the overall inequality/poverty is underestimated and the result of poverty reduction is overstated because intra-household inequality is ignored. An empirical analysis on food intake survey in the Philippines by Haddad and Kanbur (1990) show inequality in calorie adequacy could err by 30% or more comparing the individual level result and the household level result. Lise and Seitz (2011) use the collective model to estimate consumption inequality in the U.K. They find adult equivalence scale assuming no inequality among household members provides an inaccurate representation of individual consumption and underestimation of individual consumption inequality by 25% to 50%. Dunbar et al. (2013) and Bargain et al. (2014) use the conventional Rothbarth method (Rothbarth, 1943) to prove that ignoring intra-household distribution of resources leads to a large underestimation of child poverty. Kanbur (2016) mentioned when we look at the surveys that collect data only on family level, individual consumption or expenditure is calculated based on family consumption. The individual consumption is calculated by dividing family consumption by family members, sometimes equally and sometimes by other weights; -- and using equalized individual consumption would lead to underestimated total inequality. According to Kanbur the underestimation can account for 25% to 65% of the total inequality/poverty. Kanbur also claims that intra-household resource allocation influences the effect of minimum wage rise and hence intra-household inequality influences labor policies. Haddad (1992), Kanbur (1993), Hoddinott (2009) and Shi (2012) state intra-household reallocation affects the result of public programs. Public programs targeting the improvement of total welfare-- for example, school meal subsidy-- may not be able to achieve their intended goal because of intra-household transfer. Haddad and Kanbur (1993) use numerical calculation to show that considering intra-household inequality when designing the programs to target at nutritional deficiency would lead to sizable improvement of the result.

As for intra-household health inequality, Roemling and Qaim (2013) find in Indonesia, intra-household inequality rises from 1997 to 2012, especially among the overweight. Residence in an urban location, number of children, education of household head, and per capita consumption/expenditures increase intra-household nutritional inequality. The study in Coates et al. (2018) reveals significant intra-household nutrient inequities in rural Ethiopian households. And inequities are greatest for 'invisible' nutrients like microelement iron rather than protein and fat.

Sahn and Younger (2009) measure intra-household health inequality by calculating mean log deviation of BMI. They apply household survey data across 15 years in 7 developing countries and find 55% to 71% of total health inequality attributable to within household health inequality (value varies depending on the value of α in the Generalized Entropy). In their work they explain the advantages of using BMI as the indicator of well-being, including individual-level and relatively easy measurement, reflecting both food and non-food health-influencing factors, and random measure error. They reject the hypothesis of intra-household or inter-country Kuznets curve. Instead they find positive relationship between BMI inequality and wellbeing (measured in mean household BMI). On the other hand, they find no relationship between BMI inequality and per capita expenditure. This result consists with Haddad and Kanbur (1990) and Haddad et al. (1995). The study also explores the BMI ratios among family members. They find in the least advantageous families children tend to have the highest

BMI because parents may want to protect the most vulnerable family member. When family budget constraint is relaxed, the parent-to-child BMI ratio is about 1 and increases with the wealth level.

Methods

My article uses the same approach as Sahn and Younger (2009) to measure intra-household health inequality by calculating mean log deviation of BMI.

a. Inequality Measurement

I use Theil's mean log deviation, which is also the generalized entropy with $\alpha=0$ as the measurement of inequality. One advantage of mean log deviation over other inequality measurement like Gini coefficient is its decomposability, so we can divide total inequality into within group inequality and between group inequality. Population inside the same group share some identical characteristics, for example, rural-urban division groups population into rural and urban subgroups. We can estimate inequalities in groups with different characteristics and compare with overall inequality. One economic concept developed on decomposability is inequality of opportunity (Roemer, 1993), which is between group inequality (groups with different circumstance factors) while normalizing within group inequality (assuming outcome only affected by personal effort under identical circumstances in the same group).

Let n individual divide into G mutually exclusive and exhaustive groups. Let the i_{th} $(i = 1, 2, \dots n_g)$ individual in the g_{th} $(\mathrm{g} = 1, 2, \dots \mathrm{G})$ group have income or consumption $y_{i,g}$. The population share is represented by x_g and sums to 1. m_g denotes the group mean and m denotes the overall mean. If replacing $y_{i,g}$ by the group mean in each group, the inequality inside the group is suppressed into 0 and only leaves the between group inequality. Denoted inequality I, we can get within inequality $I_w = I - I_B$.

The mean log deviation can be expressed as:

$$\mathrm{L} = \frac{1}{n}\sum\sum \log\left(\frac{m}{y_{ig}}\right)$$
$$= L_W + L_B$$
$$= \sum x_g L_g + \sum x_g \log(\frac{m}{m_g})$$

where the within group MLD is the weighted sum of group MLDs.

In our case, the variable of interest is BMI rather than income, thus the province level inequality for a given province *k* can be expressed as:

$$\mathrm{I(k)} = \frac{1}{N_k}\sum \ln(\frac{\mu_k}{BMI_{i,k}})$$

where N_k represents the sample size in province *k*, μ_k stands for the mean BMI in province *k*, and $BMI_{i,k}$ means the BMI of the i_{th} person in the sample.

We can decompose provincial inequality into intra-household and between household inequality in the same way. In this case, K represents the number of households, *k*=1,2,…K, N_k stands for the number of family members in household, N expresses the total sample size, $I(k)$ is the inequality of household k, μ represents average BMI in the total sample, and μ_k is the mean BMI in the household. The first term on the right-hand-side defines within household inequality and the latter term is between household inequality. Household with $I(k)$=0 has no intra-household inequality but still contribute to total inequality through between household inequality.

$$\mathrm{I(total)} = \sum_{k=1}^{K} \frac{N_k}{N}[I(k)] + \frac{1}{N}\sum_{k=1}^{K} N_k \ln(\frac{\mu}{\mu_k})$$

b. Health Indicator

In this article I follow Sahn and Younger (2009)'s approach to use BMI as measurement of well-being. The distribution of BMI differs by age and sex for people under 20 and it has a unified standard for 20 and above. Hence, I standardize BMI for all sex and age groups by a fixed reference group, which is 19-year-old (228 month) male in our case.

$$BMI = F_{\bar{a},\bar{s}}^{-1}(F_{a,s}(bmi))$$

where F[26] is the distribution function of BMI that World Health Organization defined reference population by age and sex. *a* stands for age, *s* stands for sex and *bmi* is the actual collected BMI. In this case, $\bar{a}$=19 and $\bar{s}$=male. The standardized BMI is calculated by finding the position of the individual actual BMI in the distribution of the specific age and sex group's BMI distribution, and then find the value of that percentile in the BMI distribution for 19-year-old male. Note in the standardization, the reference group can be chosen in any sex and age. Although different age and sex group has different BMI distribution, I do not intend to calculate the absolute value of inequality. Rather, I try to find out what proportion of within group inequality attributes to the total inequality, in a relative measurement scale. Hence choosing different age/sex group does not affect the robustness of the analysis.

Sahn and Younger (2009) claims standardized BMI is an attractive representative of

[26] WHO Multicentre Growth Reference Study Group. WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development. Geneva: World Health Organization, 2006 (p229-297, p301-304).

well-being over consumption and income. It reflects and only reflects individual level well-being, and family member cannot deprive each other's BMI (at least not directly). Income and expenditure reflect both individual and household level well-being because of the arbitrary rule applied in the measurement, and expenditure usually involves public good inside a household. In addition, individual-level consumption measurement may be subject to measurement error which could be caused by survey design, interview, recall periods, and price deflation (Deaton and Grosh, 2000; Pradhan, 2001).

Alternative health indicators that reflects individual level consumption include caloric intake (Kanbur and Haddad, 1992; Haddad et al., 1995) but this indictor is potentially problematic considering the difficulty of data collection and processing. Many nutrition surveys, like CHNS, collect food intake data by either asking interviewers to recall their food consumption and or having recorders to observe and record food consumption condition. However, food leftover, false recall[27], inaccurate estimation, improper prediction may affect the accuracy of these data. In addition, comparing individual food consumption with different demographic backgrounds is unfair because people in different age, height, occupations, geographic location, health status may differ in needs for caloric intake although these could be standardized in a manner similar to the BMI data. Furthermore, BMI reflects cumulative nutrition intake in a longer term, and thus may avoid inaccuracies introduced by short-term fluctuations in food consumption. Other health measurements like happiness, activities, mortality, morbidity, life expectancy and height are infeasible for this study because they are either ordinal,

[27] del Ninno et al. (2001) find memory about food is likely to distort beyond 24-48 hours.

discrete, inaccurate, not standardized across sex and age groups, or strongly affected by other factors.

BMI is a positive, cardinal and continuous measurement. The calculation only involves height and weight which is direct and simple. The data collected on height and weight is considered accurate with random errors, unlike income whose error may be correlated to other variables. BMI reflects individual absorption of food and non-food resources like sanitary conditions and access to medical care, and it reflects consumption according to personal needs.

The major problem of using BMI is the double thresholds of the health sector. BMI value inside a specific interval is considered healthy. Either underweight or obesity indicates some health problem. Another problem of using BMI as health indicator is that it does not directly reflect illness. In addition, BMI can only reflect inequality related to partial food and health consumption, but it could not reflect other consumption inequality such as education expenditure. In our case, the sample I used in 12 Chinese provinces shows relatively low obesity rates (obesity: BMI>=30[28], 3% in sample, see the last column in Table 1) and food consumption accounts for a large proportion in total consumption in developing countries so I BMI still effectively represents a reasonable indication of well-being.

c. Kuznets Curve Estimation

Kuznets (1955) argues that aggregate inequality in a country-level population displays

[28] Division of Nutrition, Physical Activity, and Obesity, National Center for Chronic Disease Prevention and Health Promotion: If your BMI is 30.0 or higher, it falls within the obese range.

an inverted-U shape as economy develops. In the early stage of economic growth, the inequality increase, up to some point it improves with the further economic growth. Much research tests the theory across various countries and time period, and they get different results. Paukert (1973) supports the theory while Ravallion (2005) finds the relationship not convincing enough. As for health inequality, Haddad et al. (1995) apply data from Philippines to prove the common preference model generates an inverted-U relationship between unadjusted intra-household inequality of calorie intake and mean household intake, but adjusted calorie intakes display a much weaker pattern. The direct testing does not support the inverted-U relationship.

The Kuznets curve maps average income level against income inequality. In our case, it maps average household BMI against intra-household BMI inequality. Most researchers use the combination of observation and testing for the parametric estimator for the quadric function to investigate the inverted-U shape of the Kuznets curve. The non-parametric way tests inequalities in 3 different intervals: the 5^{th}-15^{th} percentile, the 45^{th}-55^{th} percentile, and 85^{th}-95^{th} percentile. If the middle interval has obvious larger inequality than two tails, then we consider the curve has an inverted-U shape. If the inverted-U shape holds, then we can infer that the intra-household BMI inequality is low when average BMI is low, so the family distribute food equally among family members in order to survive. The intra-household BMI inequality gets larger as average BMI increase, and it comes back to small disparity as average household BMI fall into health category (in a large number but smaller than 30). If the graph shows positive-U shape, it infers nutrition deprivation (large inequality) occurs in resource limited households (with low average BMI).

d. Quadratic least square regression

The visual inspection does not show a linear relationship between intra-household BMI inequality and mean household BMI (Figure 1), hence I follow Shan & Younger (2009) and assume a quadratic relationship: $y_{it} = \beta_{0it} + \beta_1 x_{it} + \beta_2 x_{it}^2 + \varepsilon_{it}$,

where y_{it} represents intra-household BMI inequality for household i in year t, x stands for mean household BMI. The sign for β_2 is positive when the model is convex and negative when the curve is concave. We can test the null hypothesis that the regression function is linear against the alternative hypothesis that it is quadratic by obtaining t value.

When using a calibration model for quantitation, the curve must be continuous, continuously differentiable and monotonic over the calibration range. To test the adequacy of OLS model, we need to test for the normality, multicollinearity, model specification, and homoscedasticity. Normality of residuals is only required for valid hypothesis testing and is not required in order to obtain unbiased estimates of the regression coefficients. There is no multi-collinearity problem in our model because we only have one independent variable. A model specification error can occur when one or more relevant variables are omitted from the model or one or more irrelevant variables are included in the model. If relevant variables are omitted from the model, the common variance they share with included variables may be wrongly attributed to those variables, and the error term is inflated. On the other hand, if irrelevant variables are included in the model, the common variance they share with included variables may be wrongly attributed to them. Model specification errors can substantially affect the estimate of regression coefficients. OLS assumes that the variance of the error term is constant (homoskedasticity). If the error terms do not have constant variance, they are said to be heteroskedastic. If this variance is not constant, then the linear regression model has heteroscedastic errors and likely to give incorrect estimates. However, heteroskedasticity does not result in biased parameter estimates. Violation of the OLS assumptions will result in overestimated confidence intervals and underestimated p-value, and thus the statistical significance level of the independent variable may also be overestimated.

Data

Data used for this study comes from China Health and Nutrition Survey (CHNS)[29] , an ongoing project that covers 12 provinces differing in geographical location, economic development and nature resource endowment in China. This project is implemented by University of North Carolina at Chapel Hill and Chinese Center for Disease Control and Prevention collaboratively. The survey uses random cluster, multistage, and weighted sampling scheme to draw country samples in each province. More than 4,400 households and 19,000 individuals are covered by the survey, with a high follow up rate. The survey is stratified into individual, family and community level, and I use

---

[29] This research uses data from China Health and Nutrition Survey (CHNS). We thank the National Institute of Nutrition and Food Safety, China Center for Disease Control and Prevention, Carolina Population Center, the University of North Carolina at Chapel Hill, the NIH (R01-HD30880, DK056350, and R01-HD38700) and the Fogarty International Center, NIH for financial support for the CHNS data collection and analysis files from 1989 to 2006 and both parties plus the China-Japan Friendship Hospital, Ministry of Health for support for CHNS 2009 and future surveys.

individual survey and household survey in this study. The survey covers topics including nutrition and physical examination, health service, social economical background, food access, farm production, income and expenditure, family relationship, time allocation at home, media exposure, etc. and each panel is linked by individual ID and household ID. Unfortunately, there are no weights[30] to adjust the data to make them representative for China due to the original survey design . Liaoning Province exits survey in 1997 when Heilongjiang Province takes its place and participates since then. Mega cities including Beijing, Shanghai and Chongqing were added in 2011. CHNS records data starting from 1989 and updates to 2015. Since most panels I use in the analysis are only updated to 2011, this study analyzed only 9 years of data: 1989, 1991, 1993, 1997, 2000, 2004, 2006, 2009 and 2011 (Table 1). Each valid entry includes variables like personal ID, family ID, survey wave, province, height, weight, and family member identifier (whether the individual is mother, father or child in the family).

[30] "Weights for CHNS", Barry M Popkin

Table III.1: Surveys, Sample sizes, Shares of overweight and obese individuals in each sample for available years from 1989 to 2011

Province	Observations	Obesity rate	Overweight rate
Beijing (11)	1279	8.37%	34.01%
Liaoning (89 91 93 00 04 06 09 11)	9278	4.34%	23.12%
Heilongjiang (97 00 04 06 09 11)	7123	3.41%	23.84%
Shanghai (11)	1466	4.91%	30.49%
Jiangsu (89 91 93 97 00 04 06 09 11)	11368	2.77%	20.66%
Shandong (89 91 93 97 00 04 06 09 11)	10258	5.14%	28.46%
Henan (89 91 93 97 00 04 06 09 11)	10814	3.84%	21.19%
Hubei (89 91 93 97 00 04 06 09 11)	10916	2.45%	15.05%
Hunan (89 91 93 97 00 04 06 09 11)	11180	2.16%	15.10%
Guangxi (89 91 93 97 00 04 06 09 11)	12930	1.09%	9.67%
Guizhou (89 91 93 97 00 04 06 09 11)	12703	1.69%	11.44%
Chongqing (11)	1291	5.42%	23.47%
Total	100605	3.00%	18.51%

Note: 25<BMI<30, overweight; BMI>=30, obese

Source: National Center for Chronic Disease Prevention and Health Promotion and CHNS 1989-2011

When calculating the adjusted BMI, I exclude individual who is pregnant or lacking limbs because these conditions can impair the measurement accuracy. I also exclude extreme values which are out of 4 standard deviation from the overall sample mean (some individuals, especially infants, leaving height or weight blank will cause extreme value). In practice, deleting the extremes has no effect on non-parametric estimates (Robinson et al., 2005).

I assume an even split of income among family members and use income per capita as another indicator of well-being in a later part of this study analyzing household member shares of BMI. In CHNS survey, the household income is calculated from direct questionnaire response about income, summation of market and non-market activities, including nonmonetary government subsidies, and reposes to questions about expenditures. The advanced method makes the income data better quality but deviates from China Statistical Yearbook. Chen et al. (2007) find income unadjusted for inflation more consist with national statistics than adjusted income. Another well-being measurement, Consumer Price Index is also available in adjusted and unadjusted form. However, China's CPI excludes housing price during calculation and considering housing occupies a large portion of household expenditure, this variable is not used to measure household economic status. Wagstaff et al. (2003) uses expenditure as the social economic status measure. This study uses income instead of expenditure because of limited reporting of expenditure data. The response rate for expenditure is low in the data set and including too many 0 entries severally impact the data sensitivity.

Findings

I find no evidence of intra-household or cross provincial Kuznets curve. I find no clear pattern of family member deprivation measured by BMI ratio, and male-to-child, female-to-child and male-to-female BMI ratio all show symmetrical distribution against y=1, hence no specific group is more advantaged in BMI measurement. Nevertheless, the intra-household inequality accounts for more than 50% of the overall BMI inequality.

a. Intra-household Kuznets curve

I map non-parametric regression for samples of all households in 12 provinces, with mean household BMI ordered from the lowest to highest on the x-axis and intra-household inequality measured in mean log deviation on the y-axis (Figure III.1). Visually the curve seems to have an inverted-U shape, and parametric regression (Table

III.3) showing negative quadratic coefficient is supportive to the observation. However, the coefficient is very small that the line is almost flat, and the R square is also very small, suggesting a weak fit. The three test points (Table III.2) tell the opposite story, which suggests in the BMI distribution, individuals in the middle range have less inequality than the extremes.

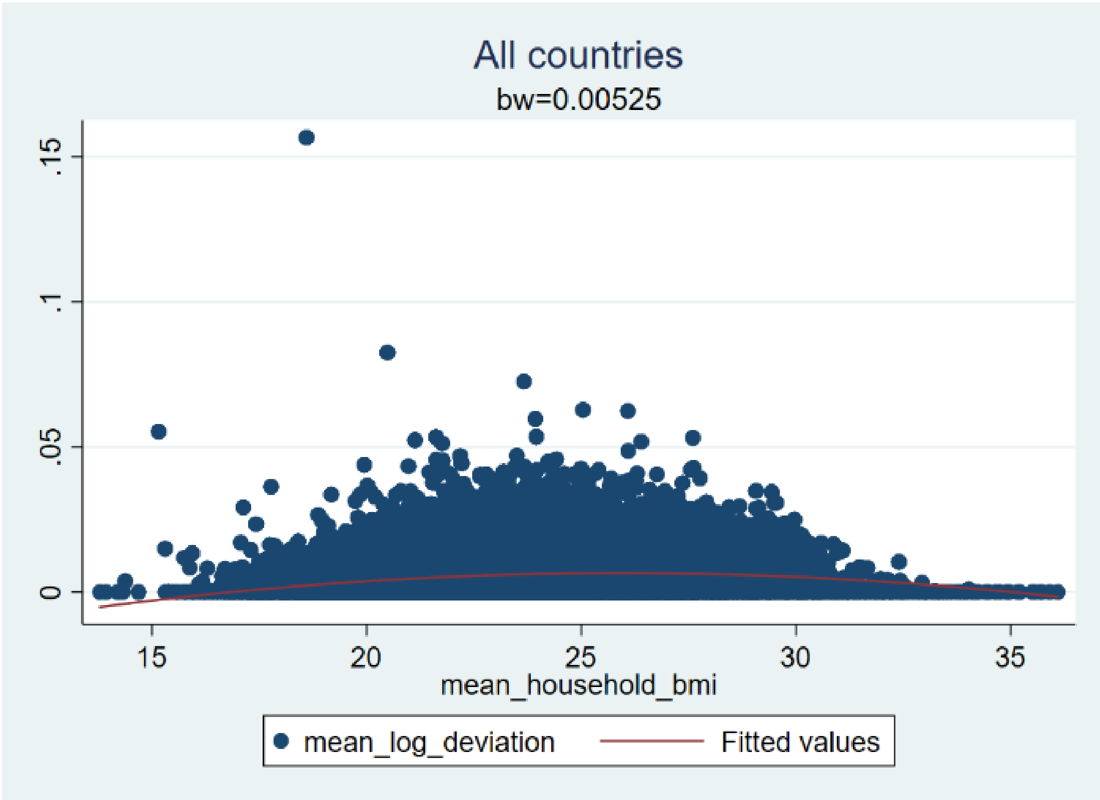


Figure III.1. Intra-household BMI inequality and household well-being, all provinces pooled, order by mean household BMI for available from 1898 to 2011.
Source: CHNS1989-2011

Table III.2. Statistical comparisons of test points of intra-household inequality for available years from 1898 to 2011.

	10th vs 50th percentile	50th vs 90th percentile
Beijing	D	I
Liaoning	D	I
Heilongjiang	D	I
Shanghai	D	I
Jiangsu	D	I
Shandong	D	I
Henan	D	I
Hubei	D	I
Hunan	D	I
Guangxi	D	I
Guizhou	D	I
Chongqing	D	I
All provinces	D	I

Note: I is an increase and D is a decrease.

Source: CHNS1989-2011 and author's calculation

Table III.3. Quadratic least square regression: relationship between intra-household inequality and mean household BMI for available years from 1898 to 2011.

VARIABLES	all	Beijing	Liaoning	Heilongjiang	Shanghai	Jiangsu
Mean household BMI	**0.00411**	**0.00385**	**0.00396**	**0.00312**	**0.00608**	**0.00394**
	[40.56]	[3.16]	[11.44]	[7.53]	[5.15]	[12.19]
Mean household BMI^2	**-7.92-e-05**	**-7.24-e-05**	**-7.63-e-05**	**-6.11e-05**	**-0.000124**	**-7.46e-05**
	[-36.57]	[-2.98]	[-10.55]	[-7.05]	[-5.10]	[-10.79]
Constant	**-0.0467**	**-0.0442**	**-0.0450**	**-0.0341**	**-0.0682**	**-0.0455**
	[-39.70]	[-2.92]	[-10.94]	[-6.91]	[-4.77]	[-12.10]
Observations	100,605	1,279	9,278	7,123	1,466	11,368
R-squared	0.04	0.013	0.028	0.014	0.018	0.043

VARIABLES	Shandong	Henan	Hubei	Hunan	Guangxi	Guizhou	Chongqing
Mean household BMI	**0.00471**	**0.00457**	**0.00472**	**0.00357**	**0.00257**	**0.00529**	-0.00172
	[14.17]	[14.21]	[17.01]	[8.92]	[7.86]	[17.24]	[-1.12]
Mean household BMI^2	**-9.15e-05**	**-8.67e-05**	**-9.21e-05**	**-6.43e-05**	**-4.41e-05**	**-0.000103**	3.99e-05
	[-13.49]	[-12.86]	[-15.36]	[-7.36]	[-5.93]	[-15.32]	[1.24]
Constant	**-0.0545**	**-0.0533**	**-0.0541**	**-0.0421**	**-0.0297**	**-0.0608**	0.0249
	[-13.44]	[-13.96]	[-16.93]	[-9.24]	[-8.29]	[-17.45]	[1.34]
Observations	10,258	10,814	10,916	11,180	12,929	12,703	1,291
R-squared	0.026	0.043	0.059	0.049	0.055	0.060	0.003

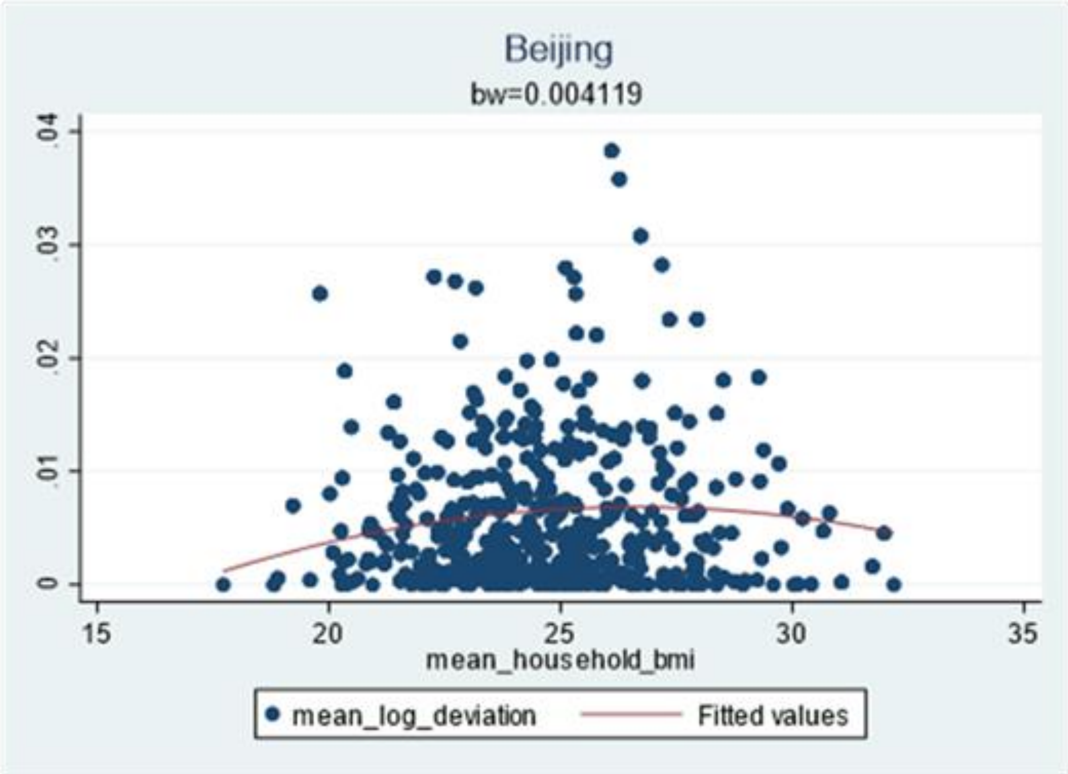
Notes: Robust t statistics in brackets; variables with bold text have 5% statistical significance level or better. Due to violation of OLS assumptions in model test, the significance level may be overestimated.
Source: CHNS1989-2011 and author's calculation based on $y_{it} = \beta_{0it} + \beta_1 x_{it} + \beta_2 x_{it}^2 + \varepsilon_{it}$

Like the overall result, whether there appears an inverted-U shape curve is undetermined for provincial results (Figure III.2). The parametric quadratic coefficients are negative and statistically significant for all provinces except for Chongqing, but still, the value is small, and the model fits poorly with a very small R square. On the contrary, the test points result demonstrates a positive U shape.

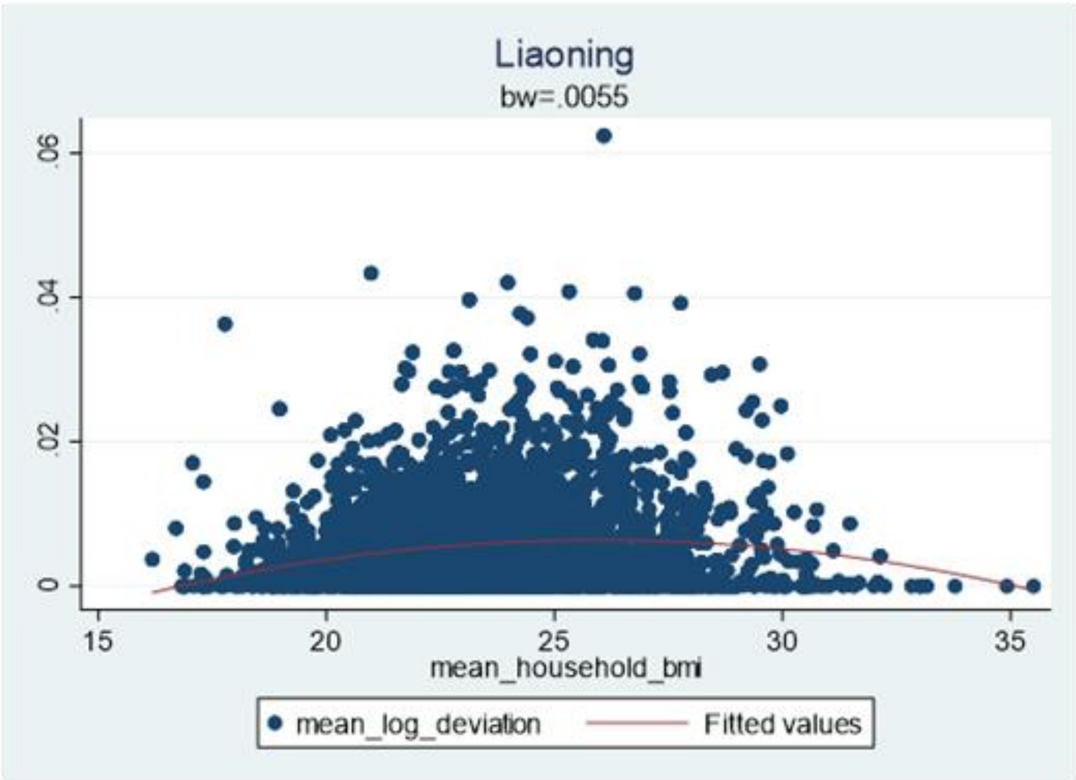
As mentioned in Methods section, we need to test the OLS model due to various reasons. Normality test (Figure III.A1 in appendix) indicates the residual is not normally distributed, thus the standard errors of OLS estimates is not perfectly reliable, which means the confidence intervals would be too wide or narrow. However, normality is not required in order to obtain unbiased estimates of the regression coefficients. The model specification test (Figure III.A2 in appendix) has a statistically significant p-value and we have to reject the null hypothesis. The result indicates that there exist omitted variables that could have better define the model. I expect model specification errors because I only have one dependent variable in the regression. The Breusch-Pagan/Cook-Weisberg test for heteroskedasticity (Figure III.A2 in appendix) shows a statistically significant p-value and we have to accept the alternative hypothesis that the residual variance is not homogenous. Cross-sectional studies are more likely to have heteroscedasticity because they often have values differ in a big range. While heteroscedasticity does not cause bias in the coefficient estimates, it does make them less precise and further from the correct population value. Heteroscedasticity increases the variance of the coefficient estimates and tends to produce p-values that are smaller than the actual value, and in turn a statistically important coefficient could be not statistically important. Thus, the statistical significance of the reported coefficients should be interpreted with caution, because the estimated confidence intervals may be too small in most cases. Overall, whether there exists a relationship between well-being measured in mean household BMI and intra-household inequality is uncertain.

b. Cross-province result

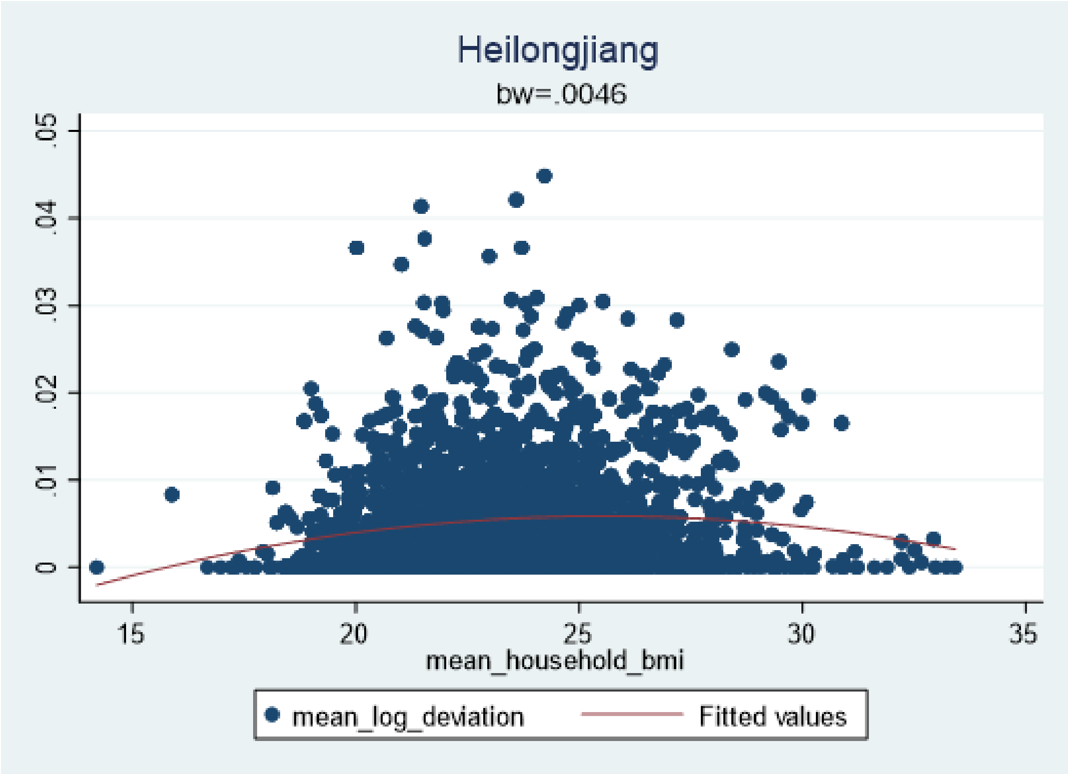
In this section I present the BMI analog to the original Kuznets curve and explore the relationship between intra-province BMI inequality and well-being indicators (provincial mean BMI and mean income per capita). I also exam whether it exits linear relationship or quadratic relationship. I plot the well-being measurement against intra-province BMI inequality calculated in mean log deviation (Figure III.3). The parametric estimation (Table III.4) shows positive linear relationship between inequality and well-being measurements, both in mean BMI and mean income, indicating inequality increase with living condition. However, both coefficient value and R square are very low, and the trend is close to a flat line, indicating very weak relationship between inequality and well-being (Figure III.3). On the other hand, the quadratic estimators are negative, but only statistically significant when it comes to income. Also the very low R square indicates very weak relationship. Because there are only 12 observations, it is hard to say if there exits any relationship between cross provincial inequality and well-being.


Beijing
bw=0.004119
.04
.03
.02
.01
0
15
20
25
30
35
mean_household_bmi
mean_log_deviation
Fitted values

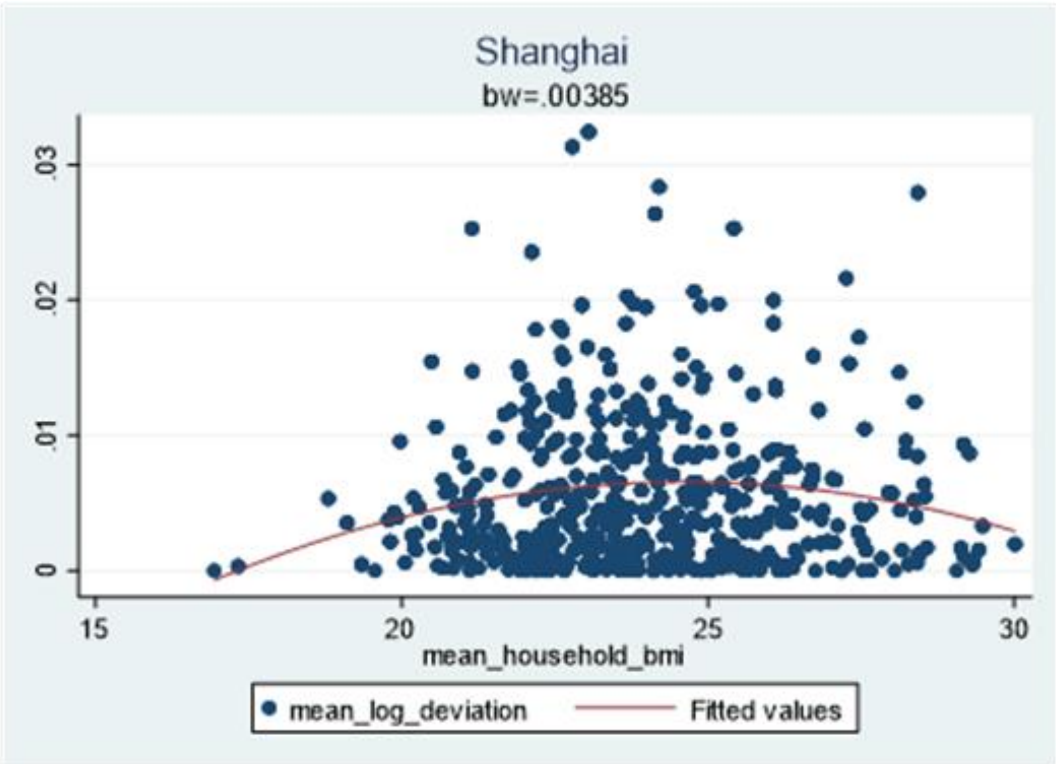
(a)


Liaoning
bw=.0055
.06
.04
.02
0
15
20
25
30
35
mean_household_bmi
mean_log_deviation
Fitted values

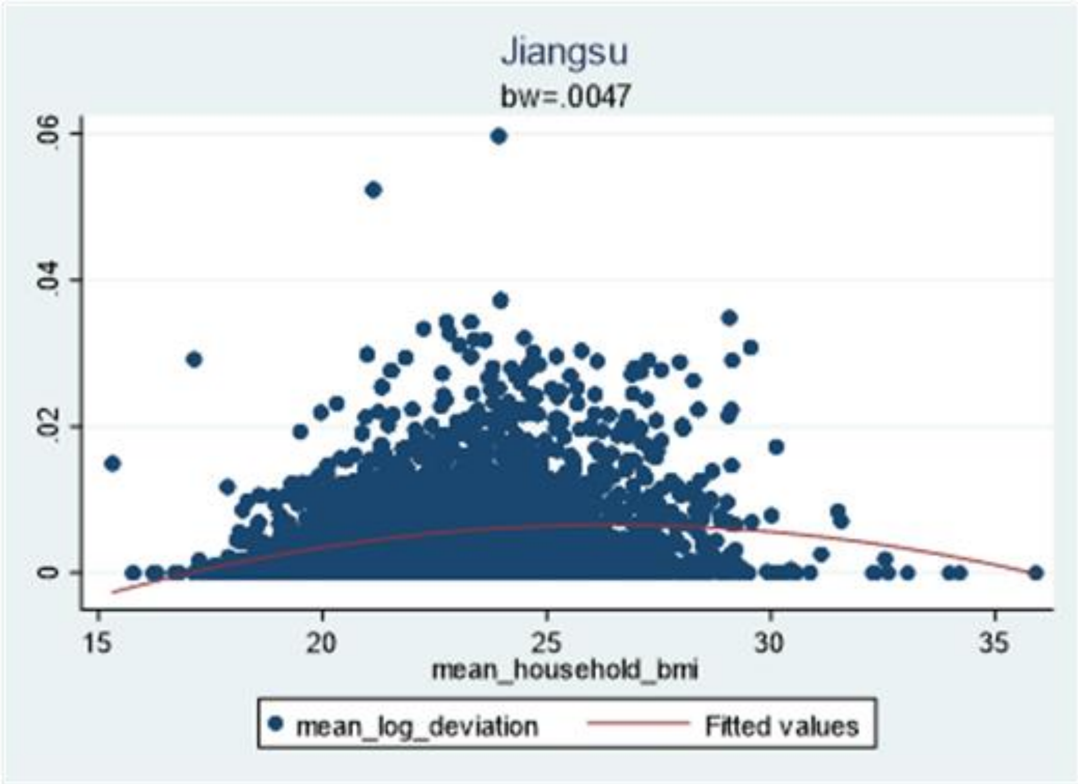
(b)


Heilongjiang
bw=.0046
0
.01
.02
.03
.04
.05
15
20
25
30
35
mean_household_bmi
mean_log_deviation
Fitted values

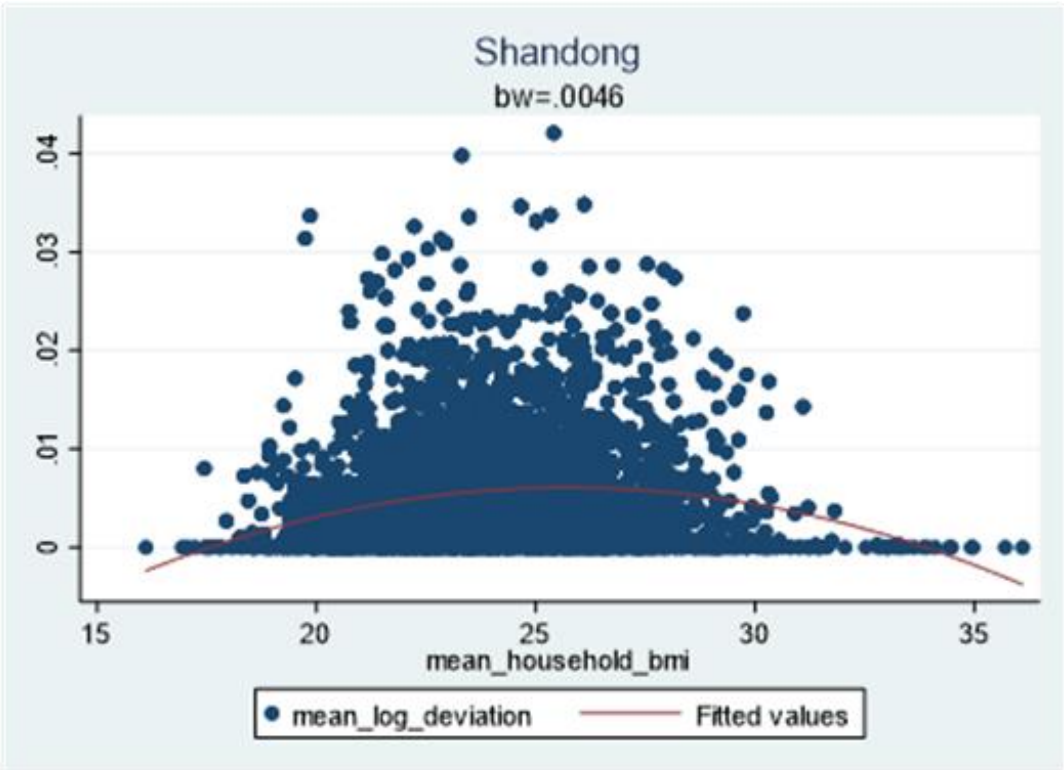
(c)


Shanghai
bw=.00385
0
.01
.02
.03
15
20
25
30
mean_household_bmi
mean_log_deviation
Fitted values

(d)

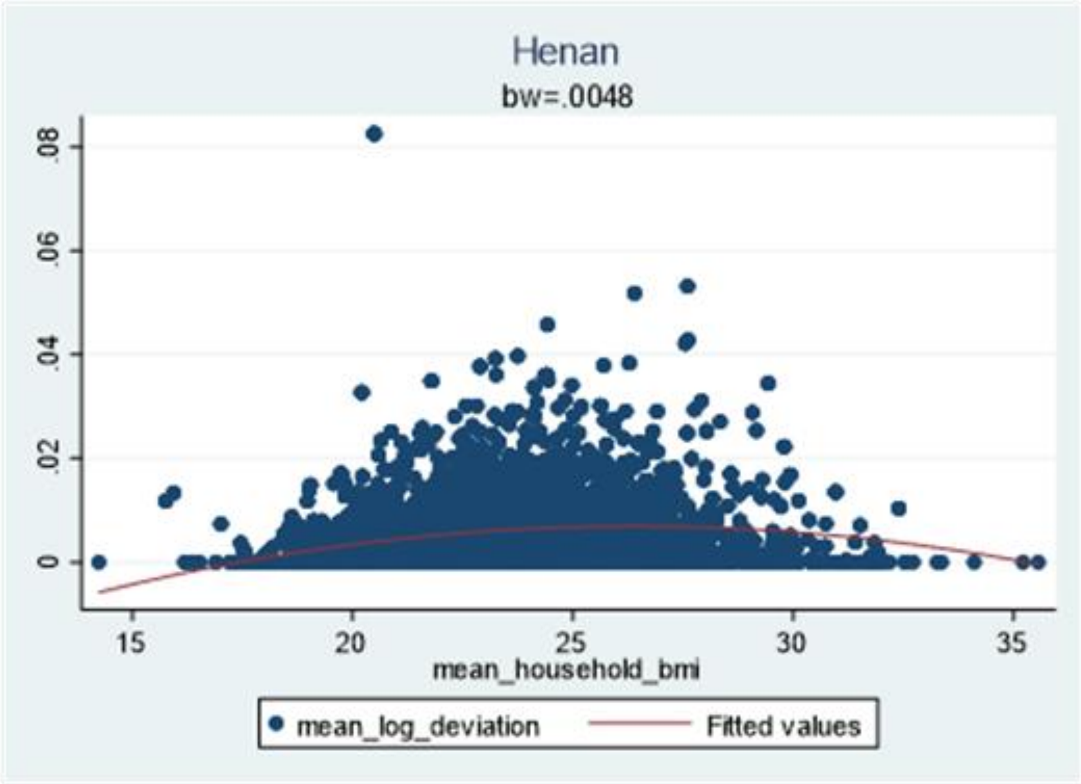

Jiangsu
bw=.0047
0
.02
.04
.06
15
20
25
30
35
mean_household_bmi
mean_log_deviation
Fitted values

(e)

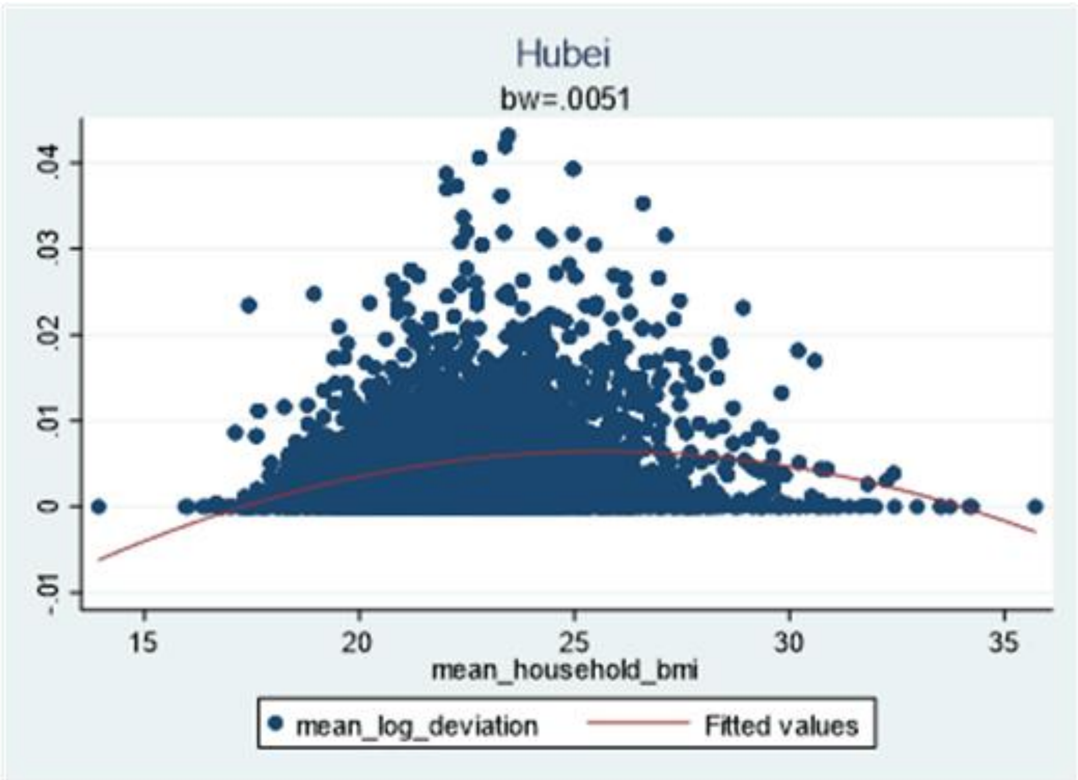

Shandong
bw=.0046
0
.01
.02
.03
.04
15
20
25
30
35
mean_household_bmi
mean_log_deviation
Fitted values

(f)

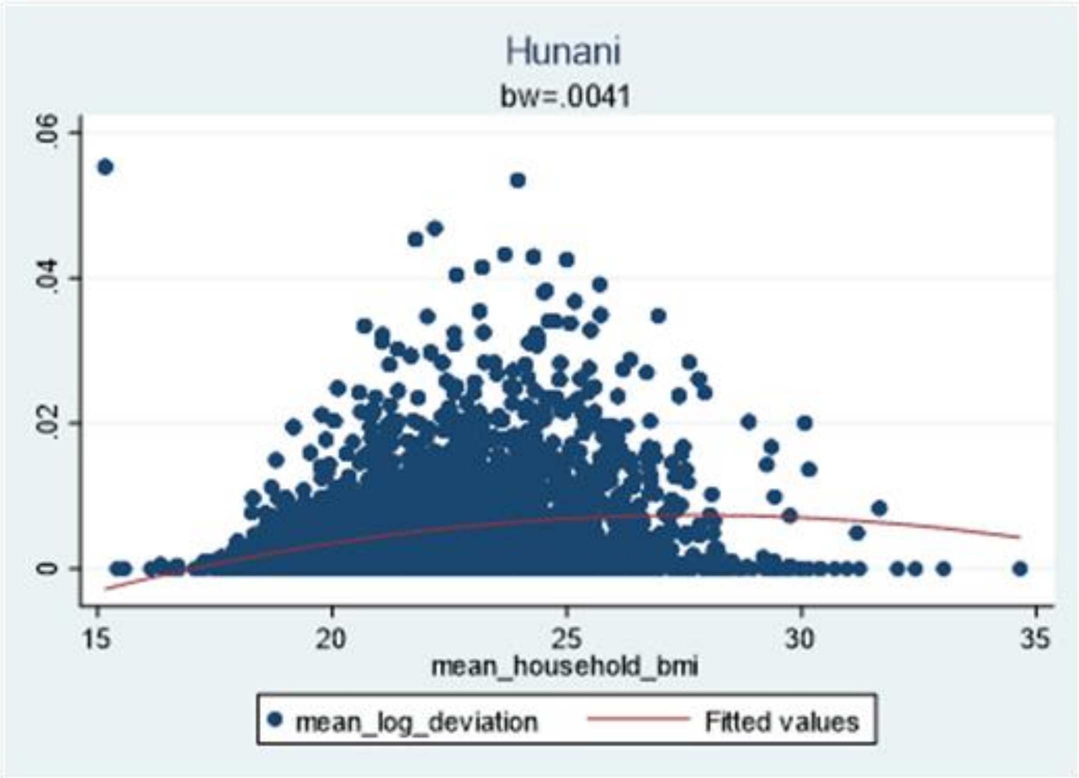
()


Henan
bw=.0048
.08
.06
.04
.02
0
15
20
25
30
35
mean_household_bmi
mean_log_deviation
Fitted values

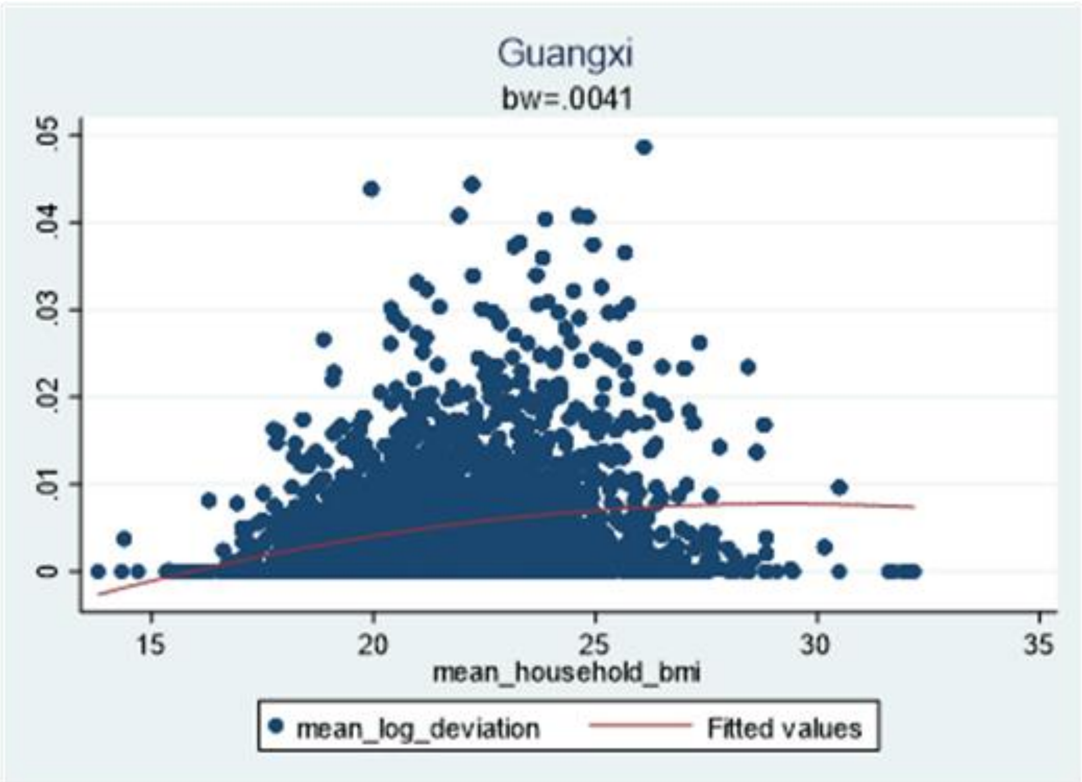
(g)


Hubei
bw=.0051
.04
.03
.02
.01
0
-.01
15
20
25
30
35
mean_household_bmi
mean_log_deviation
Fitted values

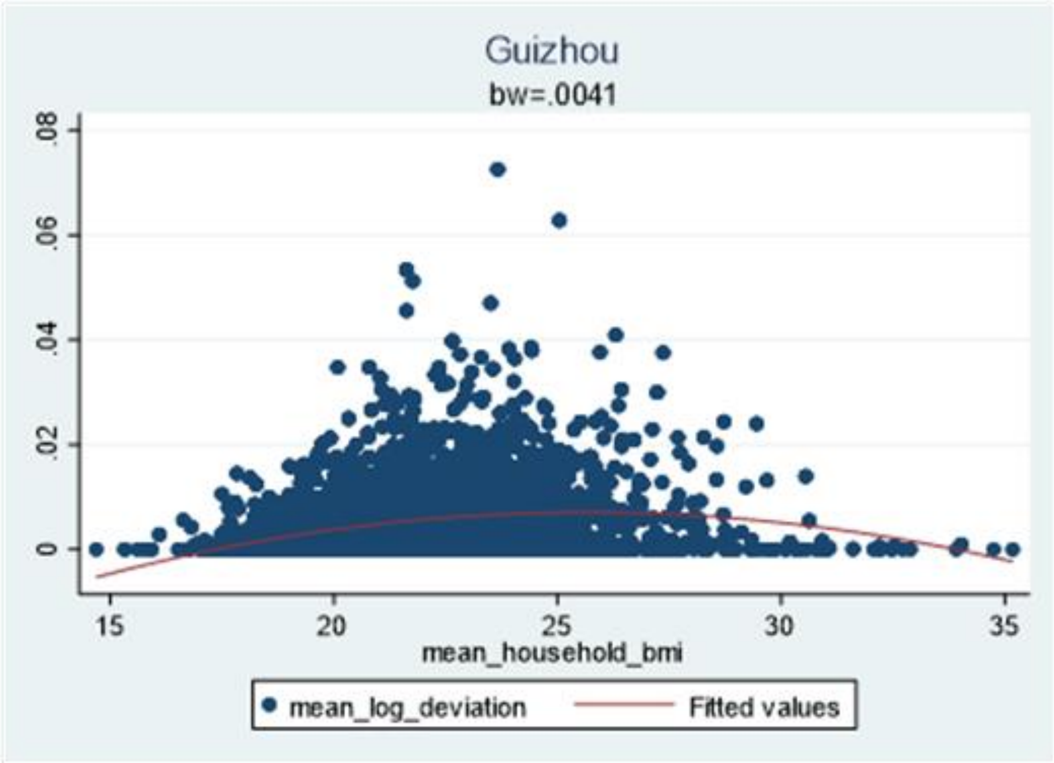
(h)


Hunani
bw=.0041
.06
.04
.02
0
15
20
25
30
35
mean_household_bmi
mean_log_deviation
Fitted values

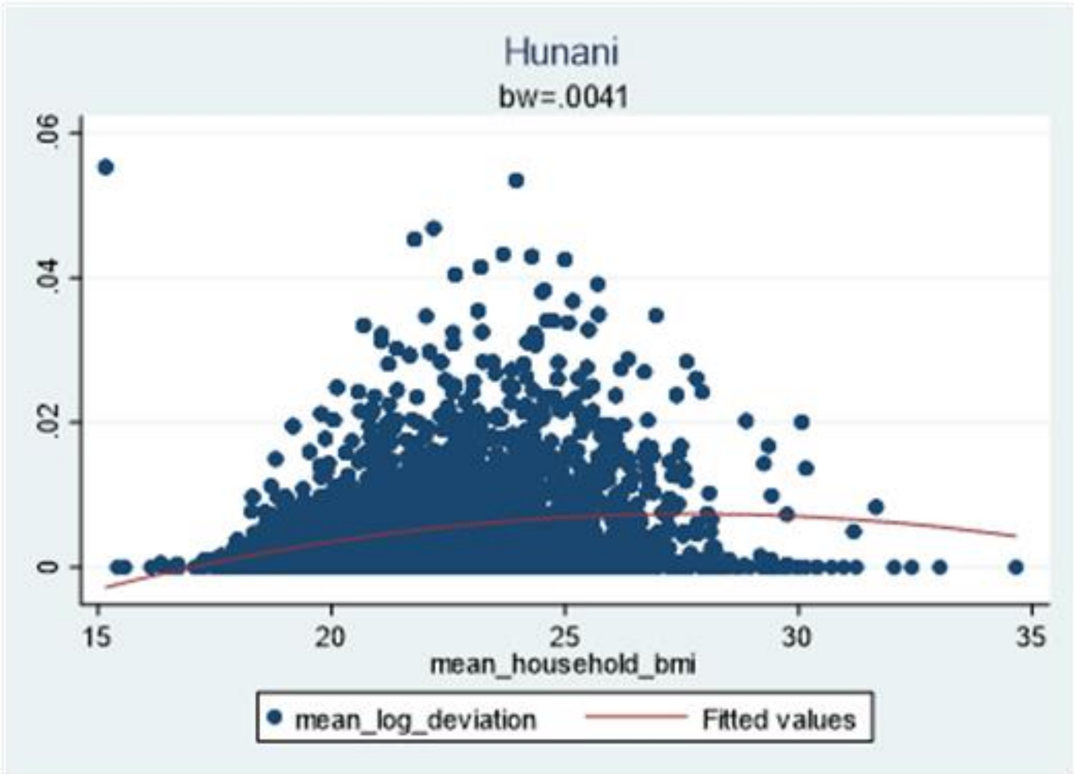
(i)


Guangxi
bw=.0041
.05
.04
.03
.02
.01
0
15
20
25
30
35
mean_household_bmi
mean_log_deviation
Fitted values

(j)



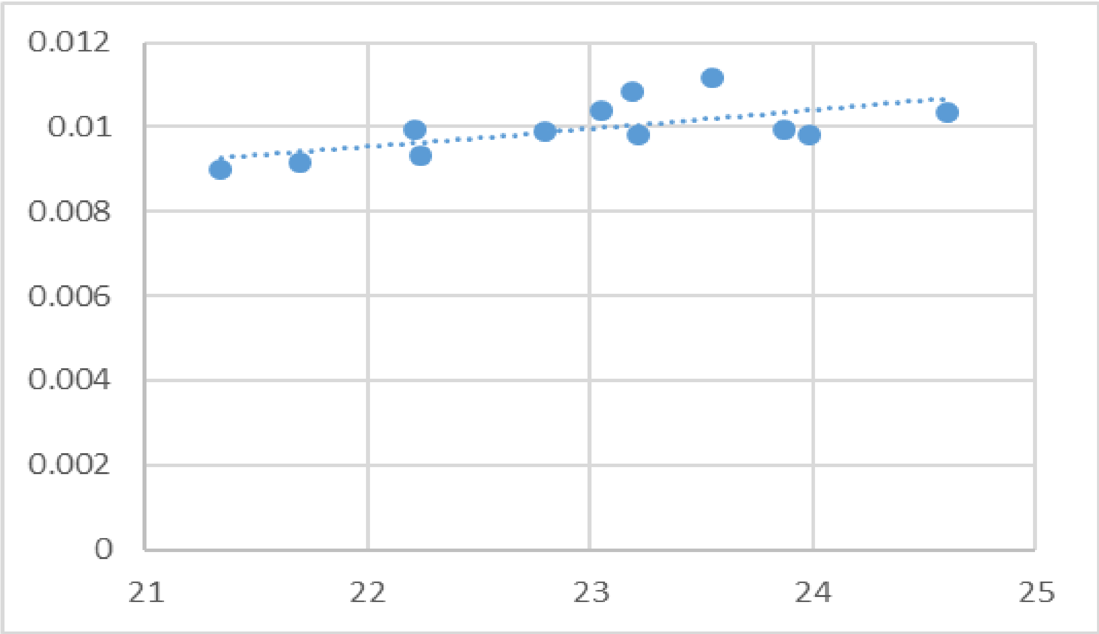
(k)



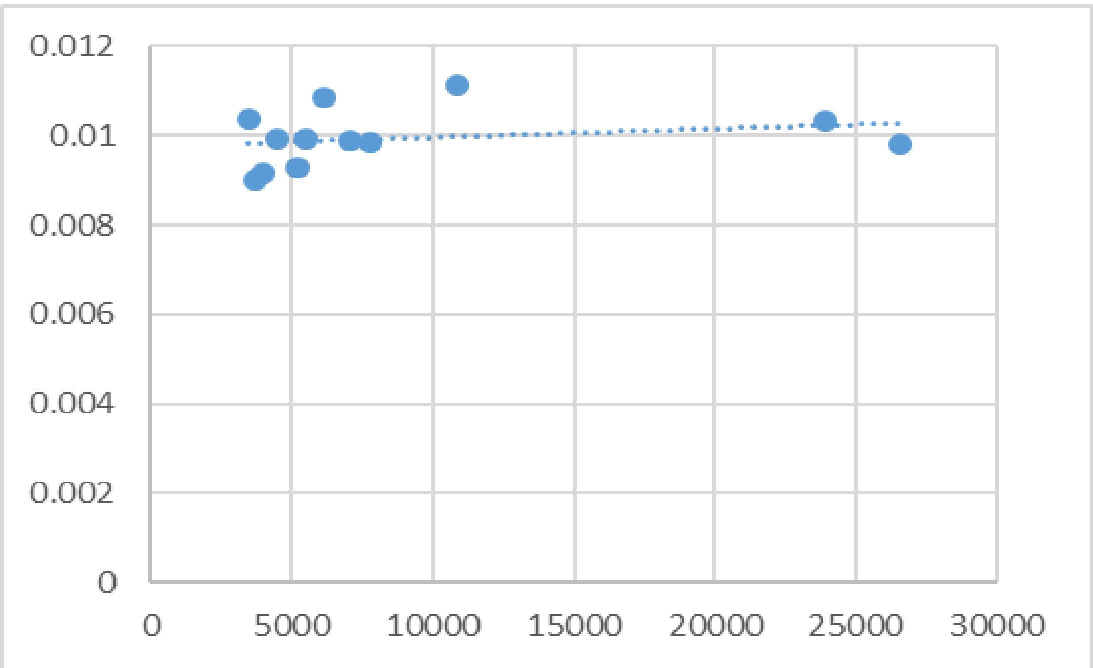
(l)

Figure III.2 (a)~(l). Provincial intra-household BMI inequality by mean BMI for available years from 1898 to 2011.

Source: CHNS 1989-2011



(a)



(b)

Figure III.3. Cross-provincial relationship between BMI inequality and well-being represented by (a) mean provincial BMI and (b) mean income per capita for available years from 1898 to 2011.
Source: CHNS 1989-2011

Table III.4. Cross province regression: relationship between intra-household BMI and well-being for available from 1898 to 2011.

	Quadratic	Linear	Quadratic	Linear
Mean BMI	0.0122	**0.0004**		
	[1.79]	[2.63]		
Mean BMI squared	-0.0003			
	[-1.73]			
Mean income per capita			**3.24E-07**	1.83E-08
			[2.27]	[0.72]
Mean income per capita square			-1.03E-11	
			[-2.17]	
Constant	-0.134	0.0002	**0.0085**	**0.01**
	[-1.72]	[0.05]	[12.58]	[32.84]
Observations	12	12	12	12
R-squared	0.56	0.41	0.24	0.05

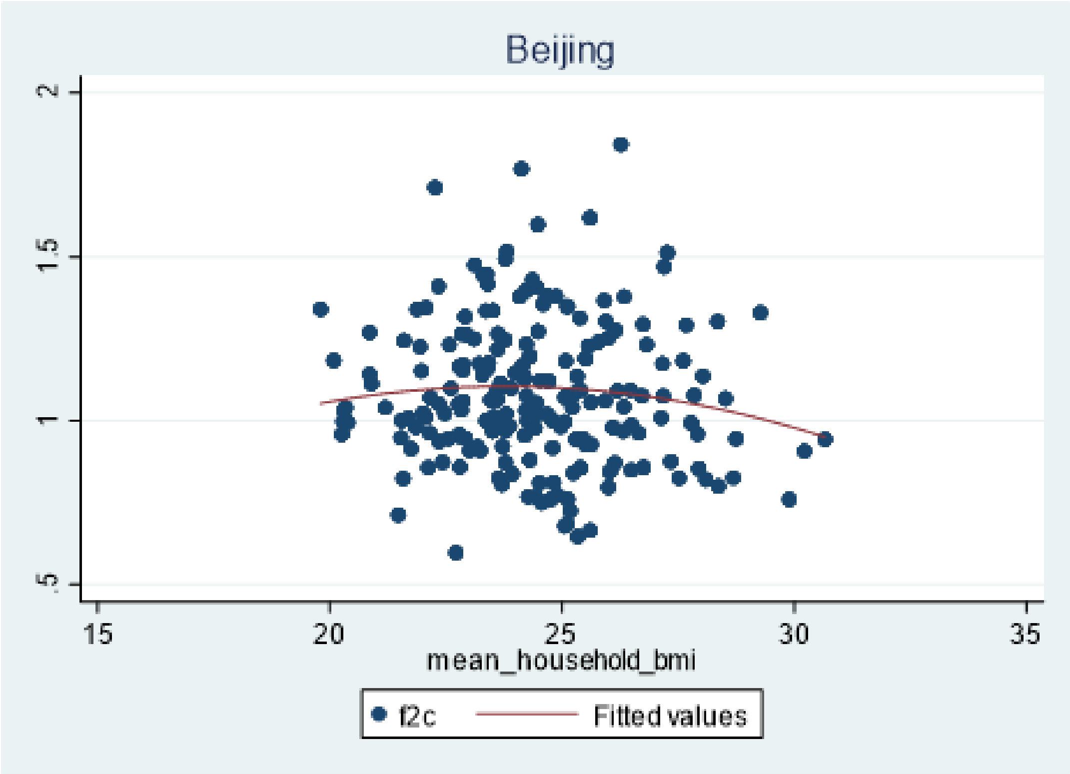
Notes: t statistics in brackets; variables with bold text have 5% statistical significance level or better. Due to violation of OLS assumptions in model test, the significance level may be overestimated.
Source: CHNS 1989-2011

c. Household member shares of BMI

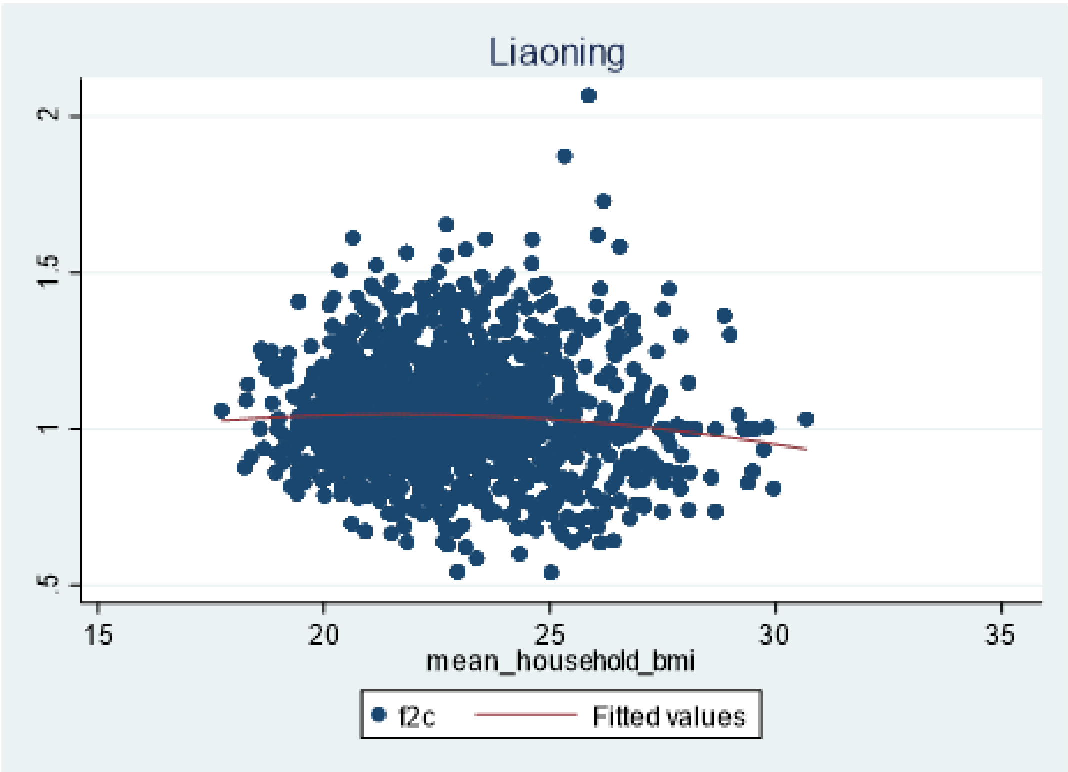
In order to access whether families will preferentially allocate health- related resources to relatively more vulnerable family members, I compare the BMI ratios among family members, ordering families from the least to the most advantageous in well-being, represented by mean household BMI and income per capita separately. I conclude there is almost no differences among family member's BMI.

Sahn and Younger (2009) find declining and flat (fe)male-to-child BMI ratios and households with low BMIs usually have ratios above 1 while households with high BMIs have ratios above 1. Their result suggests adults allocate health related resources to children in low level well-being status. Once the resource limitation is relaxed adults will gain more weight. However, in our research, the relationship between mean well-being level and adult-to-child BMI ratio is not clear. The result display almost flat and slightly declining father-to-child and mother-to-child BMI ratio as mean household BMI increases(Figures III.4 and III.5). The data points assemble into a cluster rather than displaying a clear linear relationship. The graphs are basically symmetrical around a ratio of 1. This result is consistent with Haddad et al. (2003) and Ssewanyana et al. (2008)'s findings that nutritional outcomes only has modest correlation with household resources in developing countries.

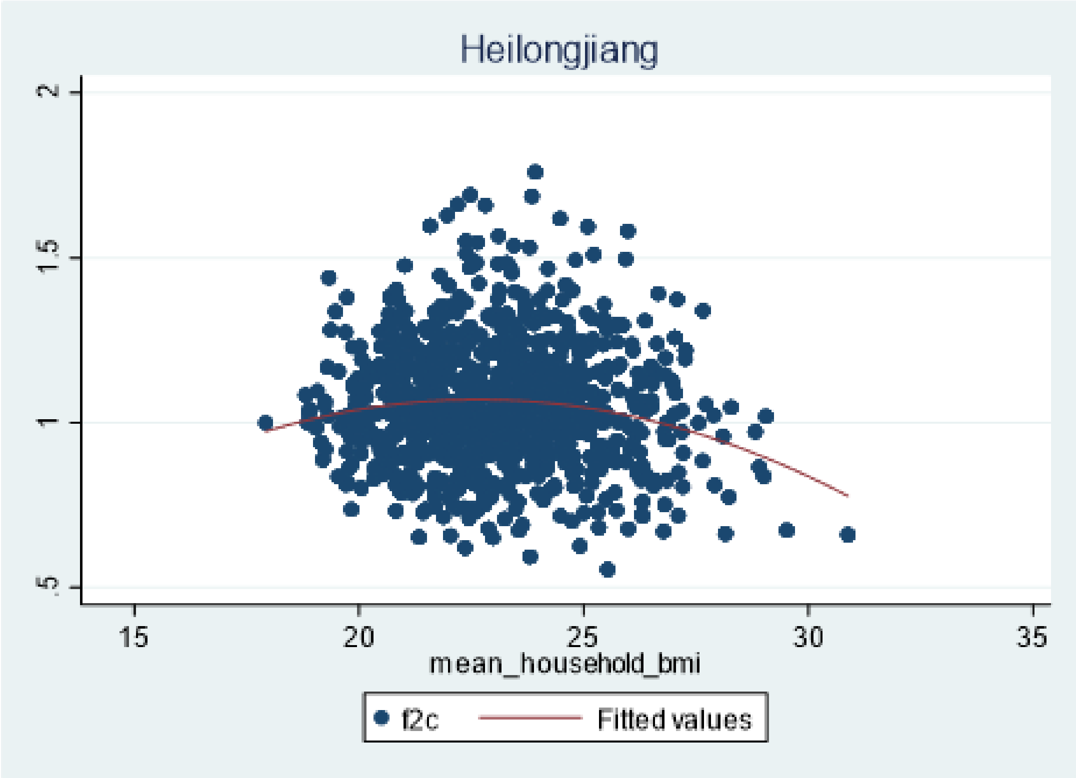
Also of interest is how BMI differences among family members change as economic status improves, whether resource-constrained families protect or deprive children and women. Sahn and Younger (2009) find declining and flat relative BMI for children as expenditure per capita increases in less developed countries. They also note in the household with lowest expenditure children have the highest BMI relative to adult male and female because adults try to protect the most venerable young children with limited resources. In China, flat and slightly decreasing father-to-child and mother-to-child BMI ratio exist, which distribute almost symmetrically against y=1 line (Figures III.6 and III.7). Amongst the families with the lowest income, both high ratio which suggests deprivation and low ratio indicating resource allocation towards children occur, in half and half.


Beijing
2
1.5
1
.5
15
20
25
30
35
mean_household_bmi
f2c
Fitted values

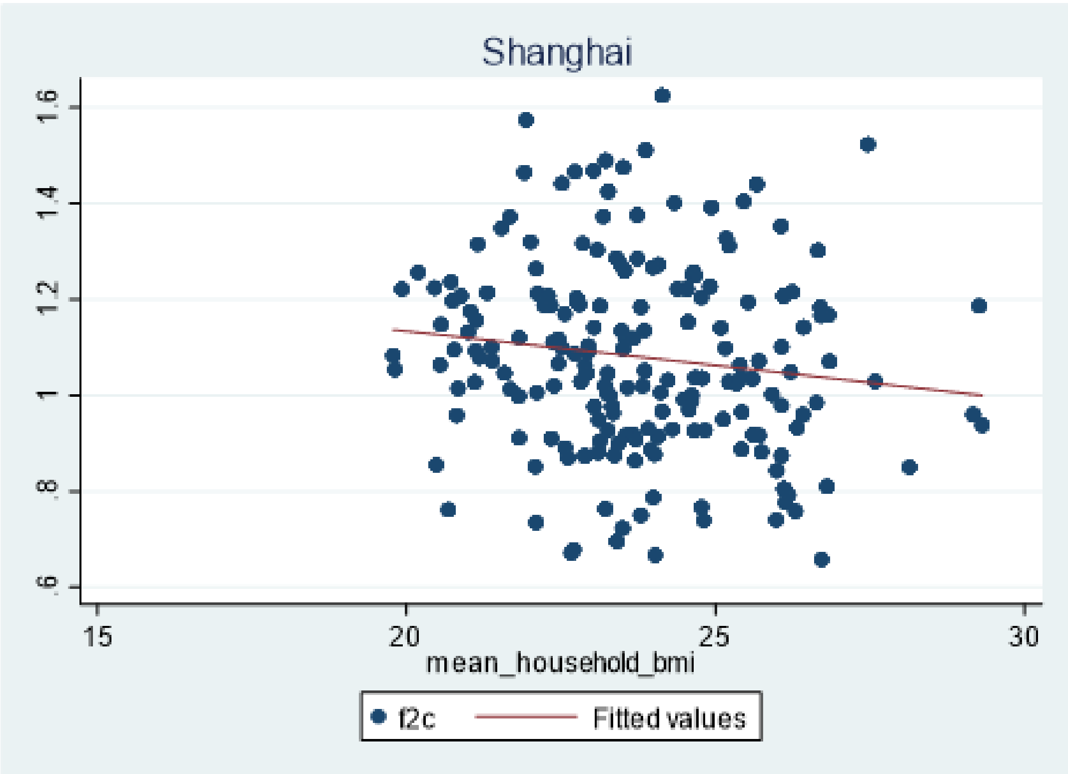
(a)


Liaoning
2
1.5
1
.5
15
20
25
30
35
mean_household_bmi
f2c
Fitted values

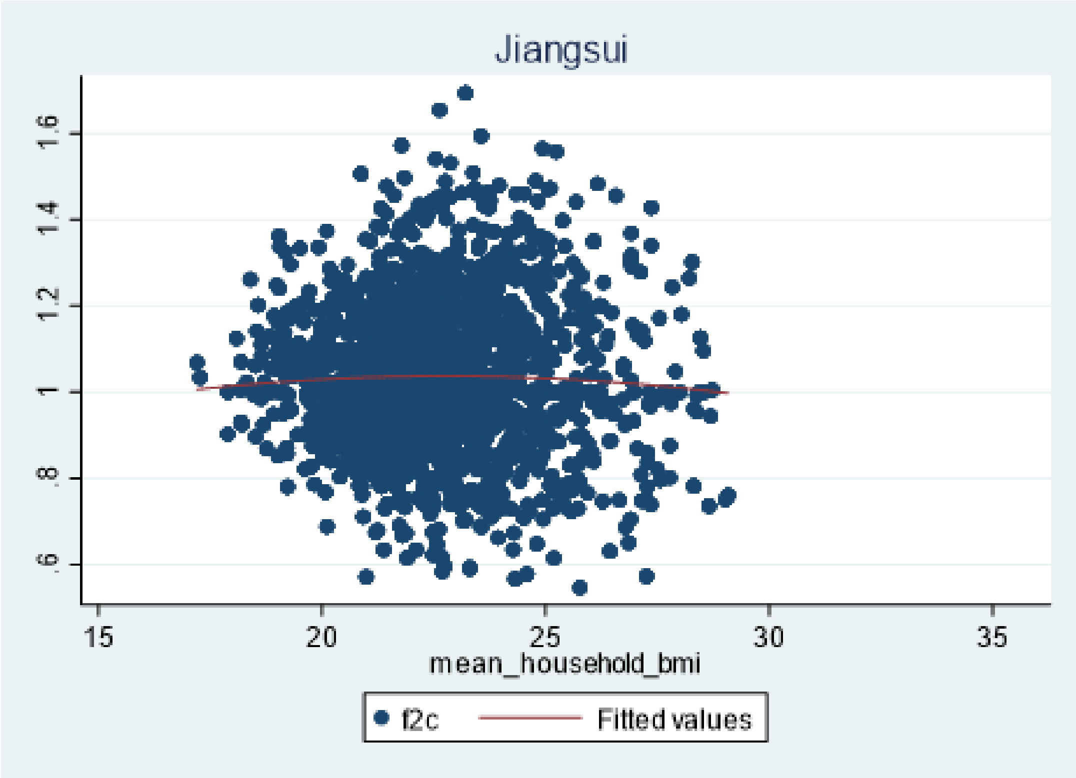
(b)


Heilongjiang
2
1.5
1
.5
15
20
25
30
35
mean_household_bmi
f2c
Fitted values

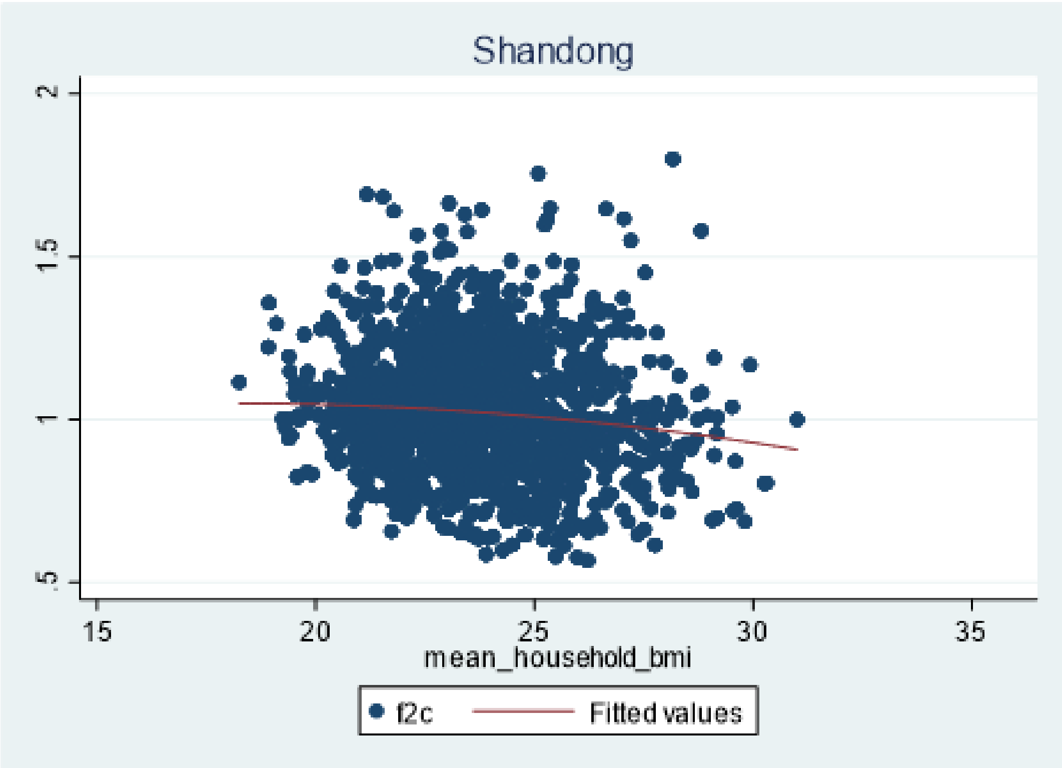
(c)


Shanghai
1.6
1.4
1.2
1
.8
.6
15
20
25
30
mean_household_bmi
f2c
Fitted values

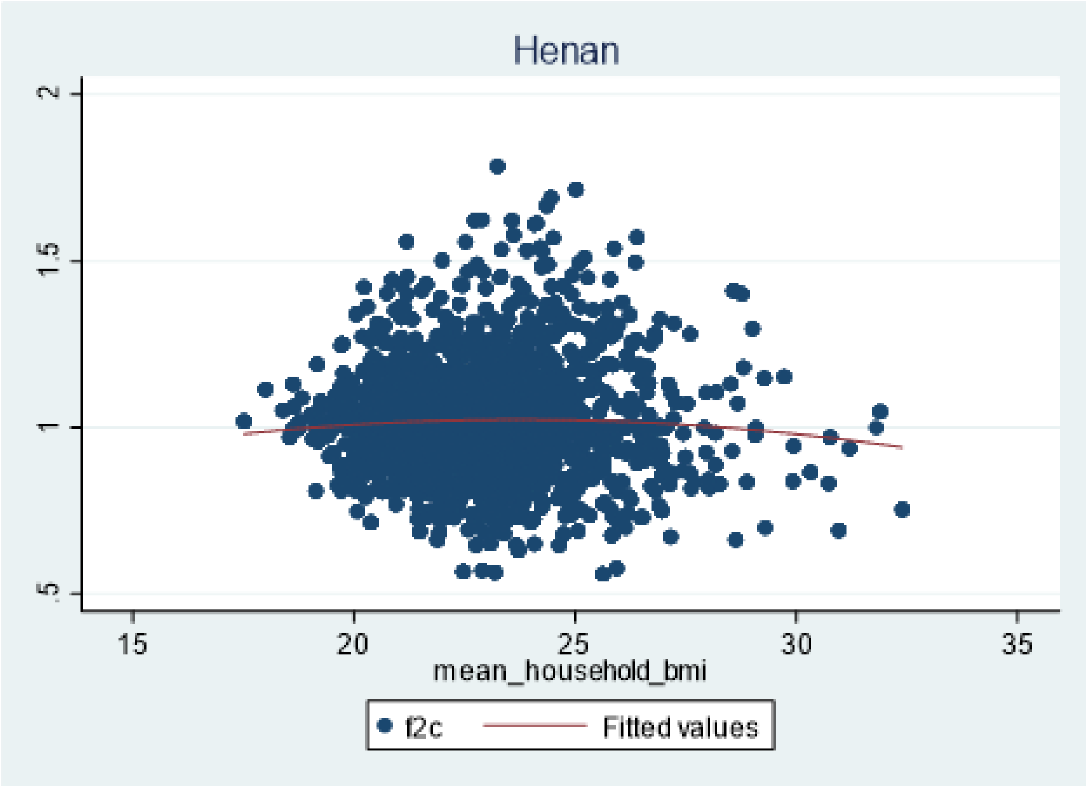
(d)


Jiangsui
1.6
1.4
1.2
1
.8
.6
15
20
25
30
35
mean_household_bmi
f2c
Fitted values

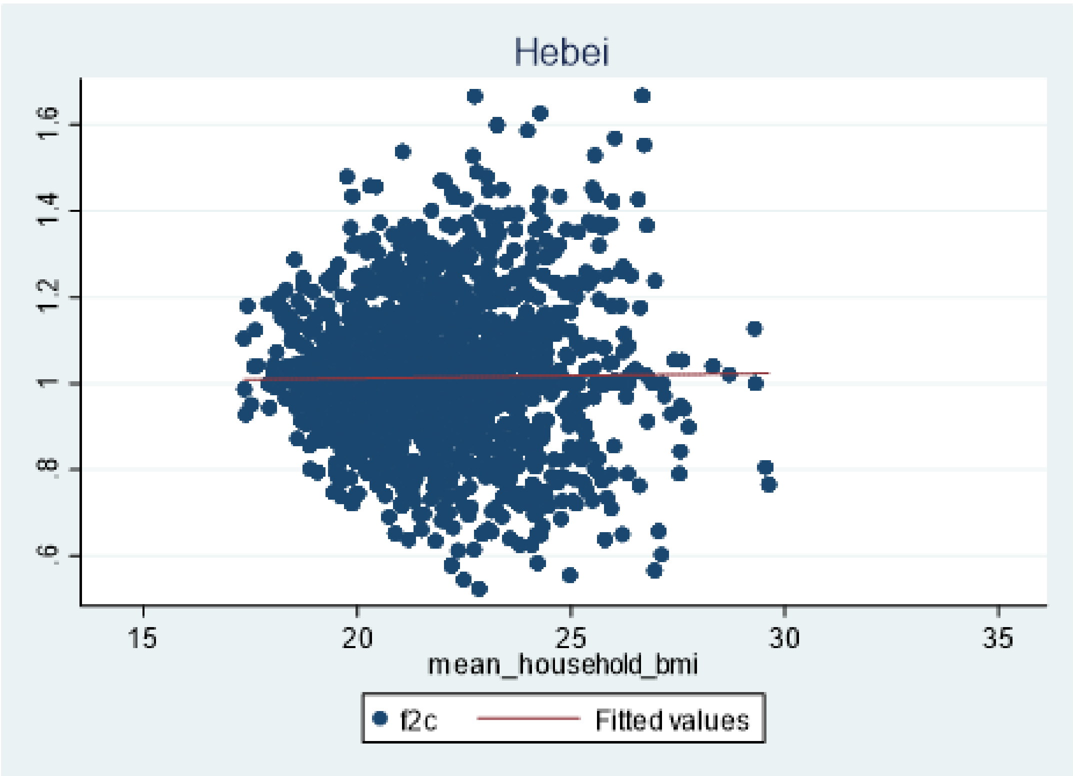
(e)


Shandong
2
1.5
1
.5
15
20
25
30
35
mean_household_bmi
f2c
Fitted values

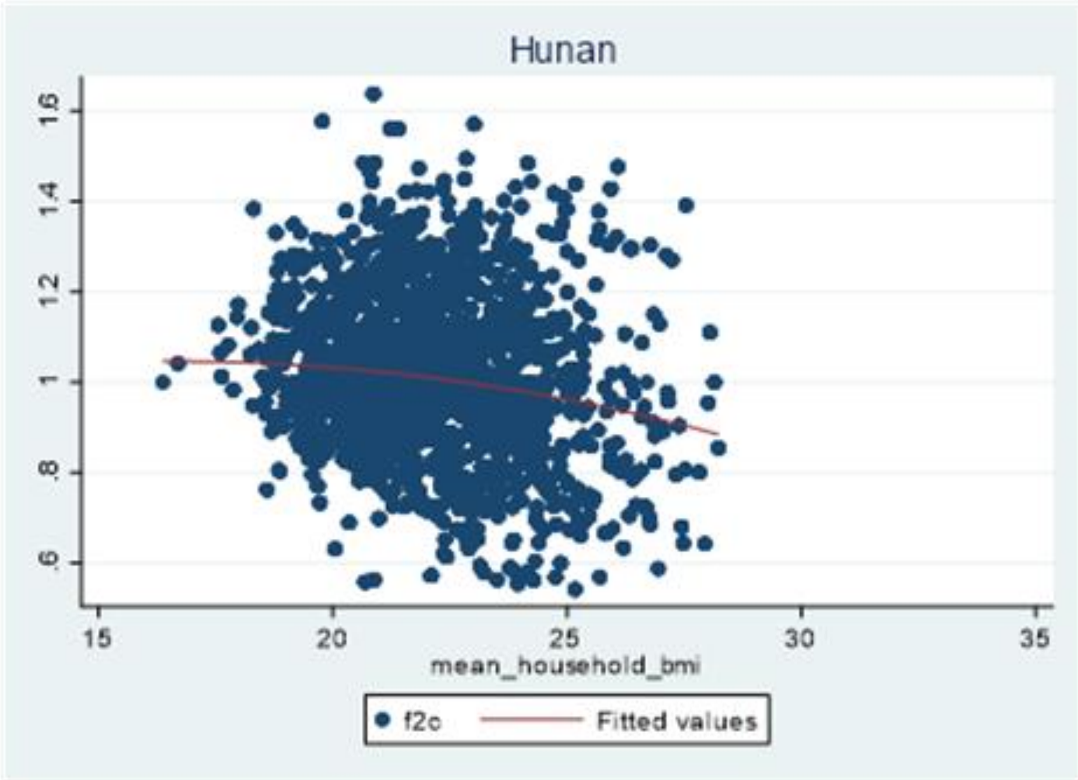
(f)


Henan
2
1.5
1
.5
15
20
25
30
35
mean_household_bmi
f2c
Fitted values

(g)

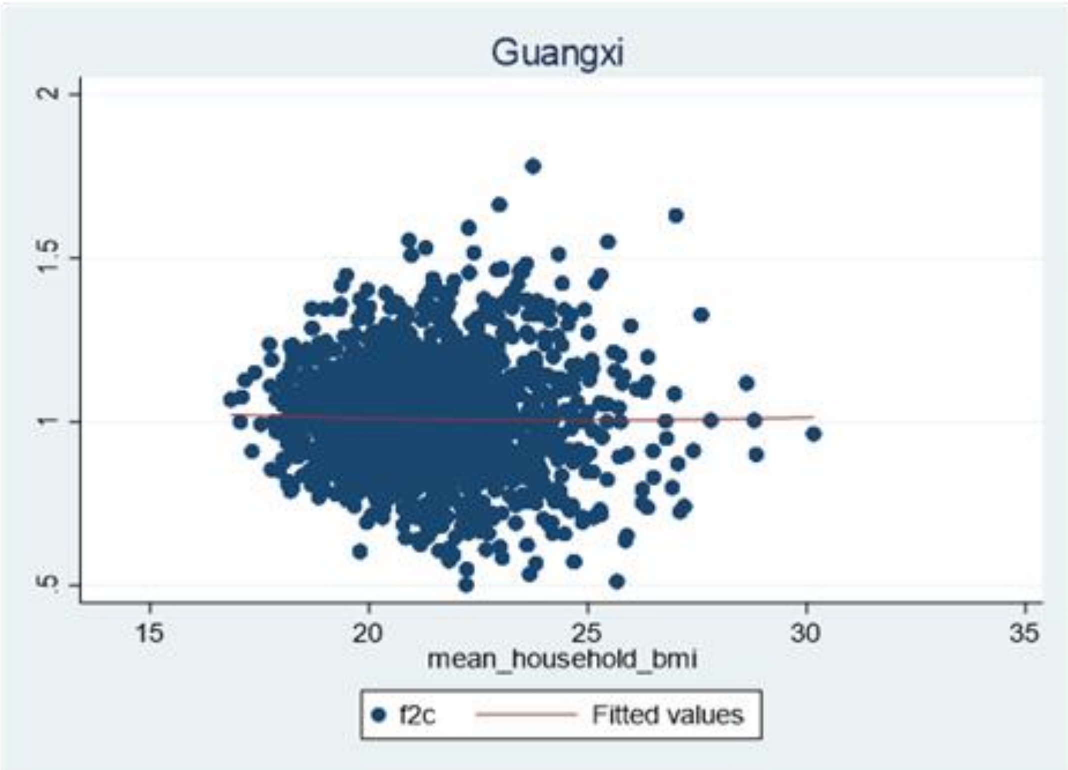

Hebei
1.6
1.4
1.2
1
.8
.6
15
20
25
30
35
mean_household_bmi
f2c
Fitted values

(h)



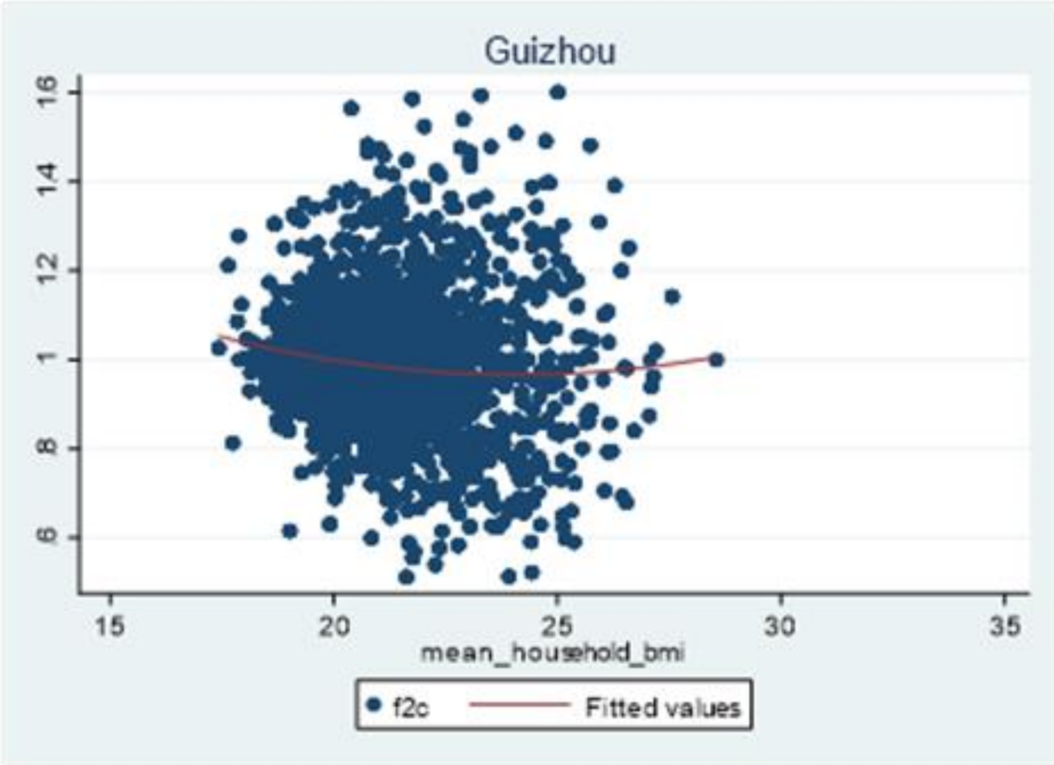
Hunan
1.6
1.4
1.2
1
.8
.6
15
20
25
30
35
mean_household_bmi
f2c
Fitted values

(i)

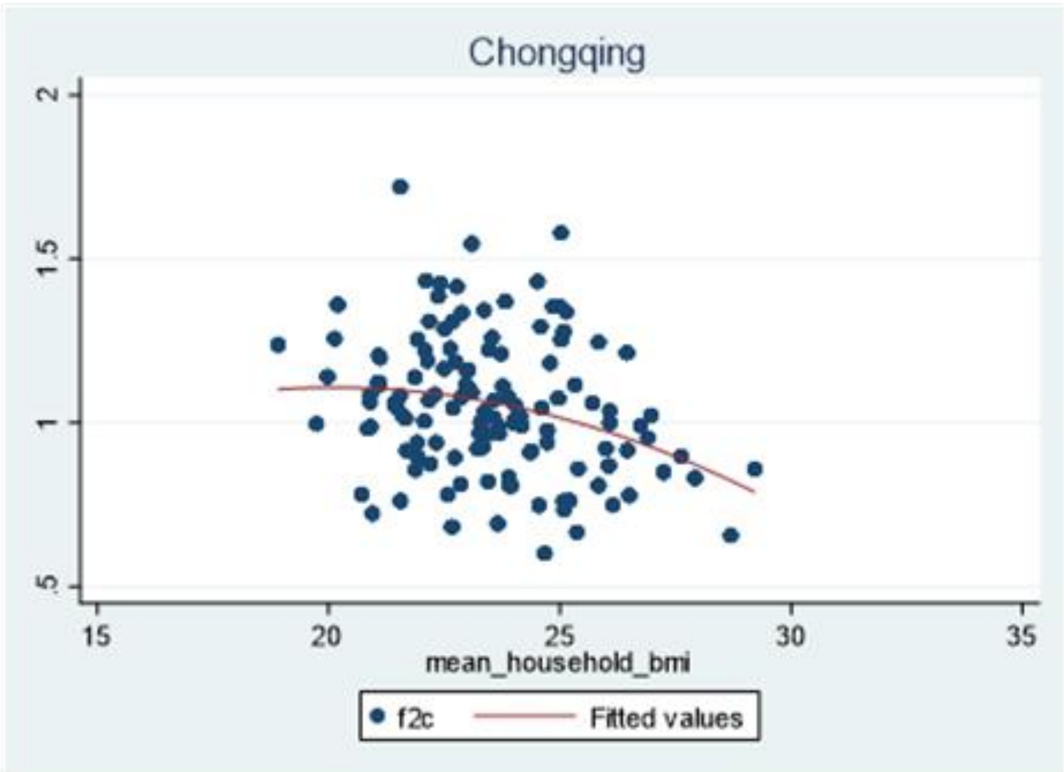


Guangxi
2
1.5
1
.5
15
20
25
30
35
mean_household_bmi
f2c
Fitted values

(j)

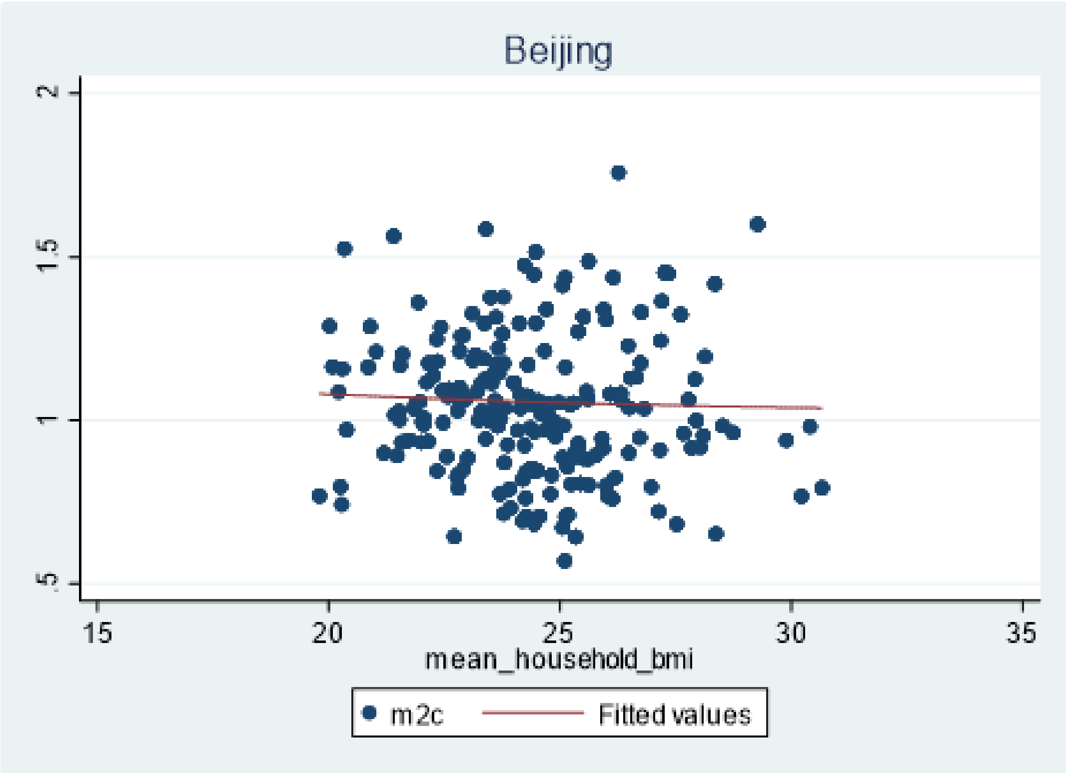


(k)

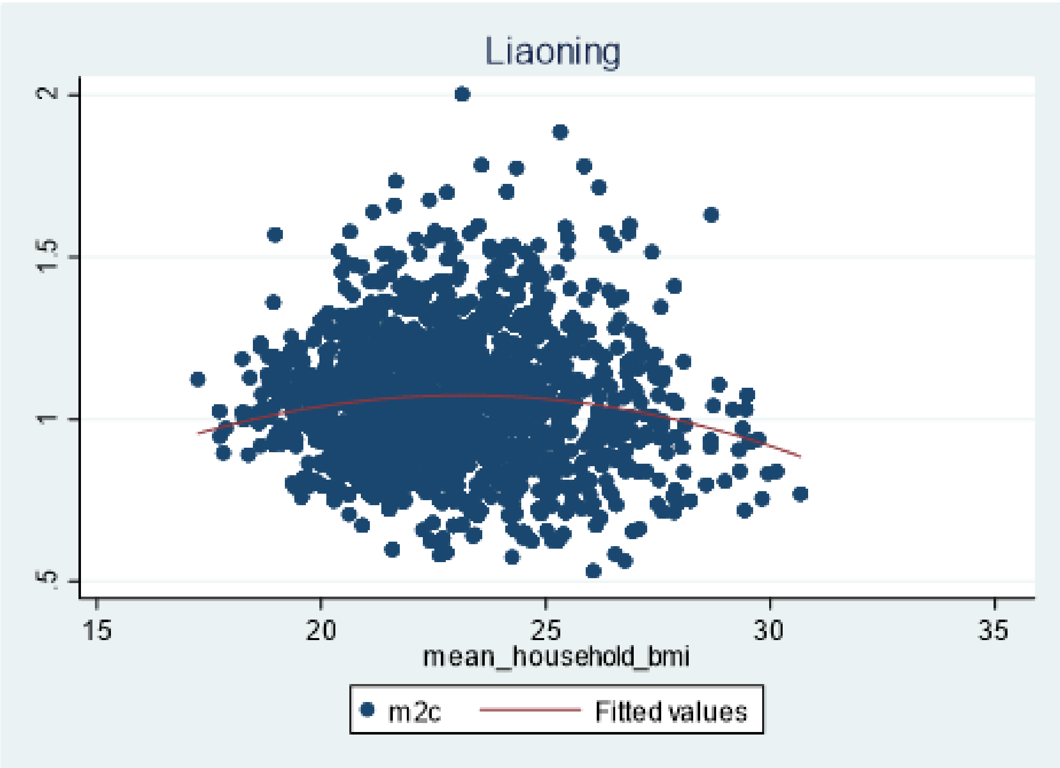


(l)

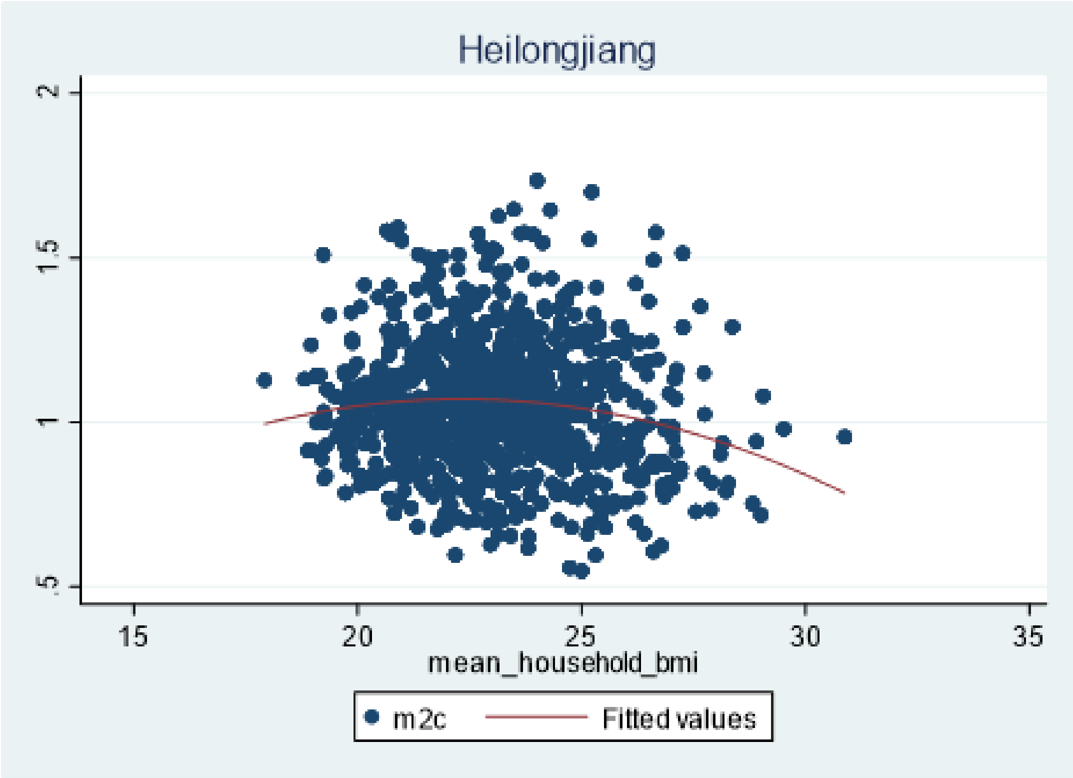
Figure III.4 (a)~(l). Father-to-child BMI ratio ordered by mean household BMI for available years from 1898 to 2011.
Source: CHNS1989-2011


Beijing
2
1.5
1
.5
15
20
25
30
35
mean_household_bmi
m2c
Fitted values

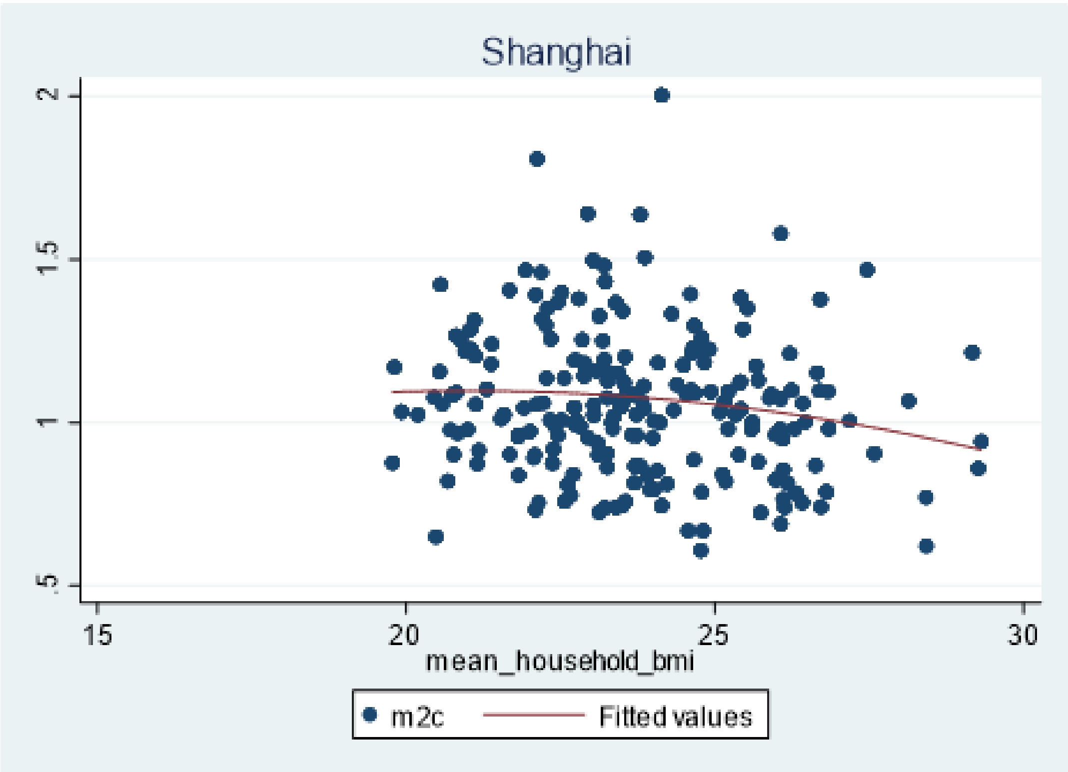
(a)


Liaoning
2
1.5
1
.5
15
20
25
30
35
mean_household_bmi
m2c
Fitted values

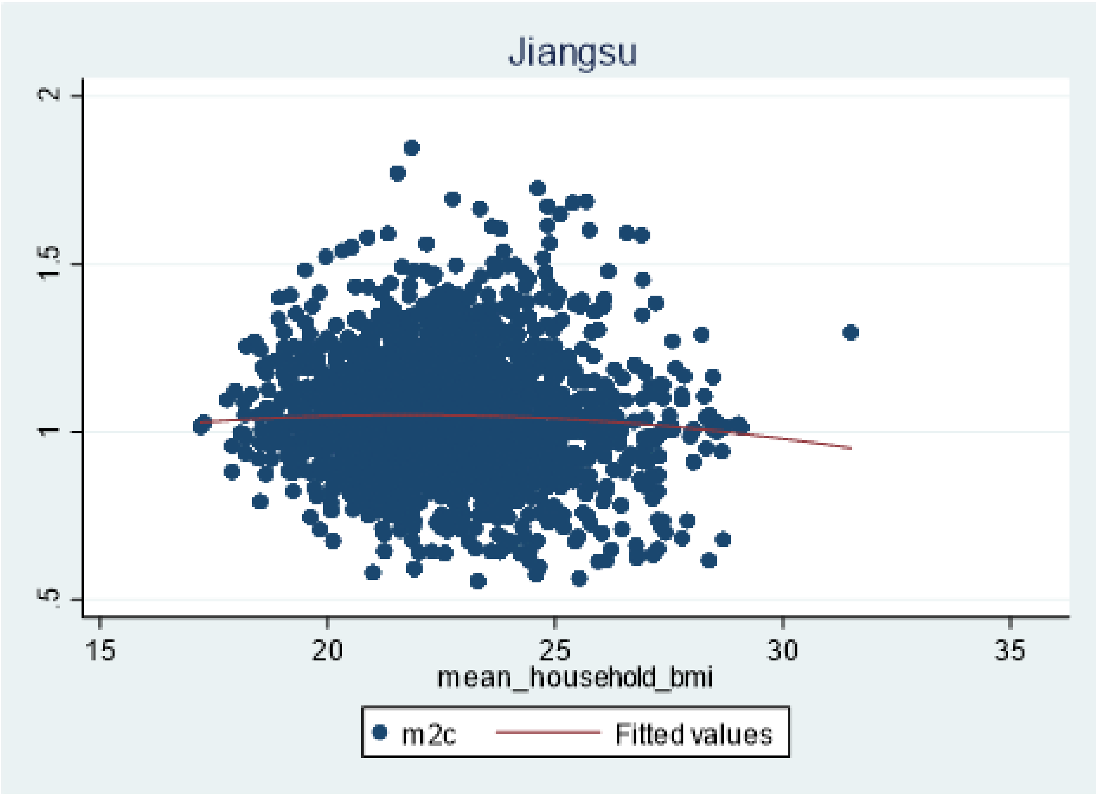
(b)


Heilongjiang
2
1.5
1
.5
15
20
25
30
35
mean_household_bmi
m2c
Fitted values

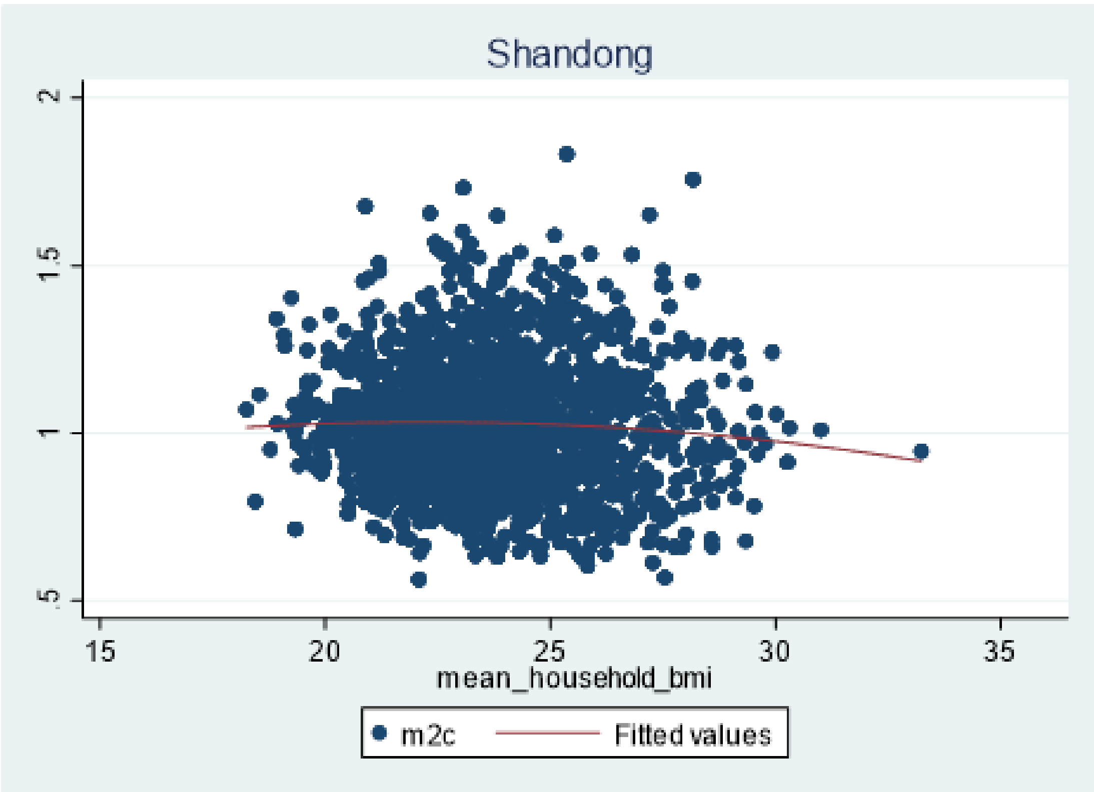
(c)


Shanghai
2
1.5
1
.5
15
20
25
30
mean_household_bmi
m2c
Fitted values

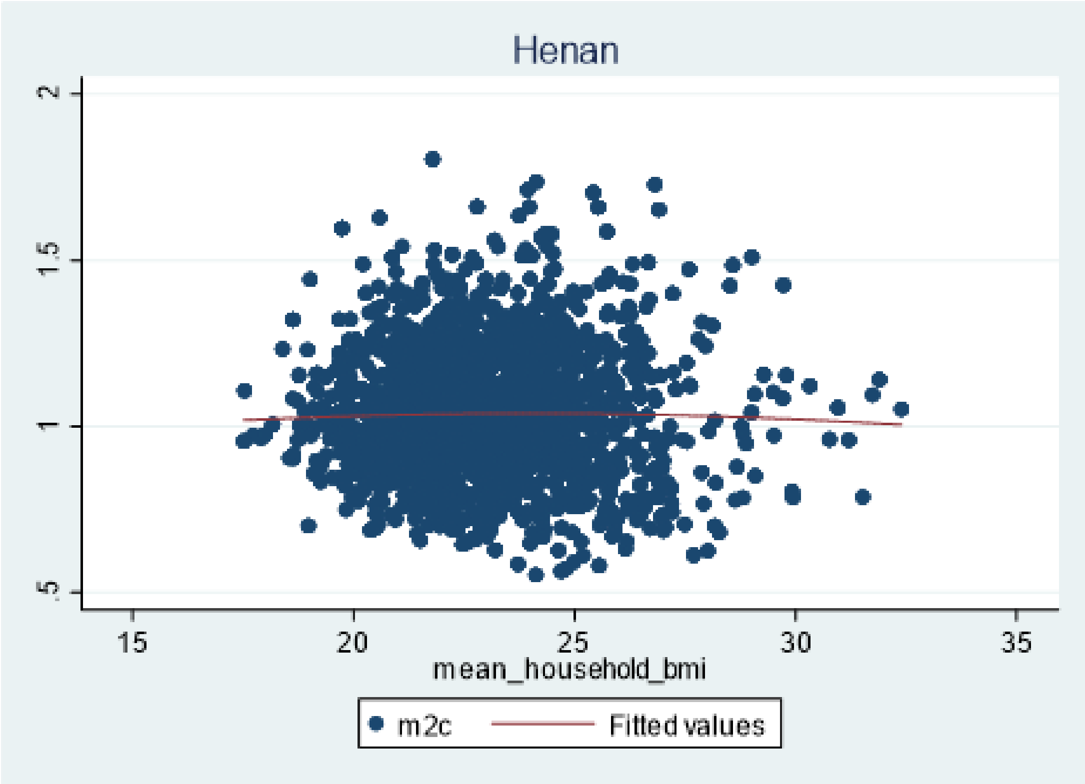
(d)


Jiangsu
2
1.5
1
.5
15
20
25
30
35
mean_household_bmi
m2c
Fitted values

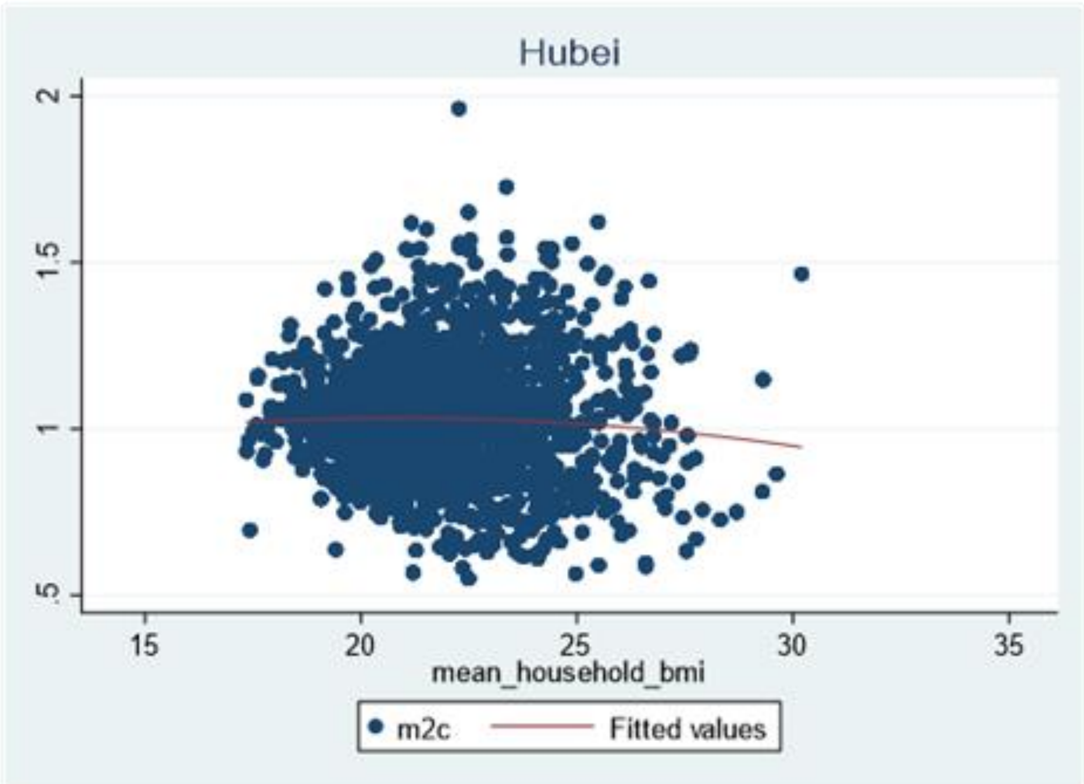
(e)


Shandong
2
1.5
1
.5
15
20
25
30
35
mean_household_bmi
m2c
Fitted values

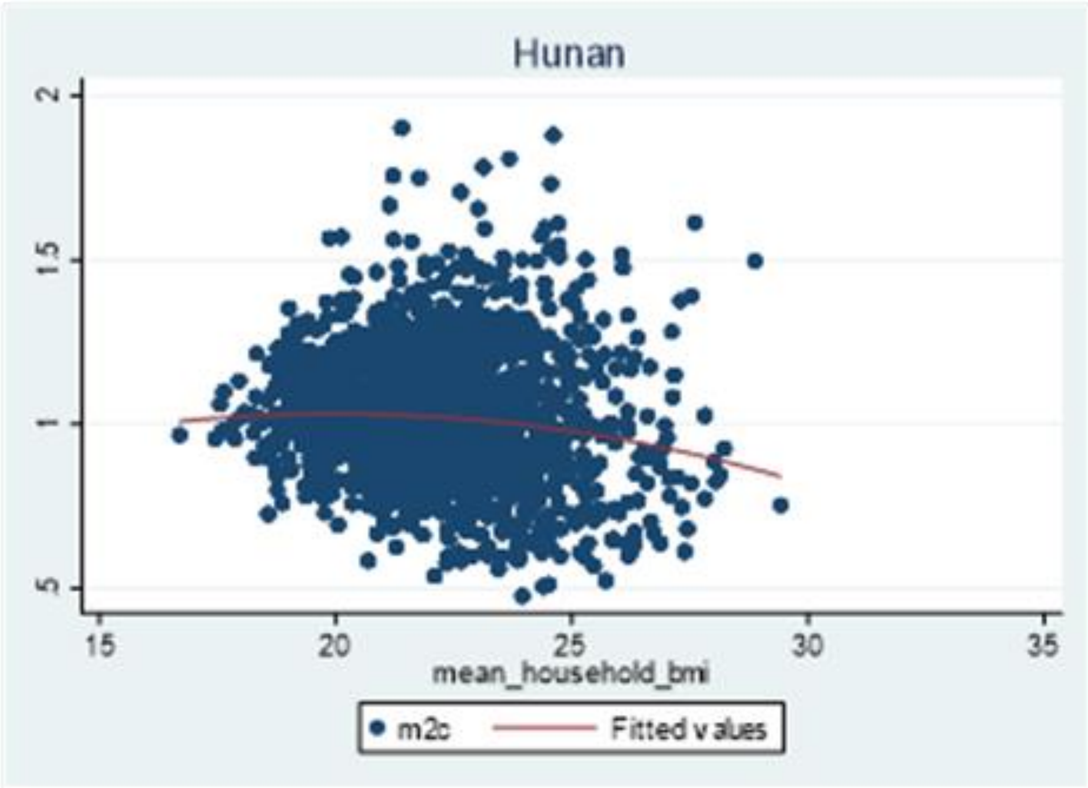
(f)


Henan
2
1.5
1
.5
15
20
25
30
35
mean_household_bmi
m2c
Fitted values

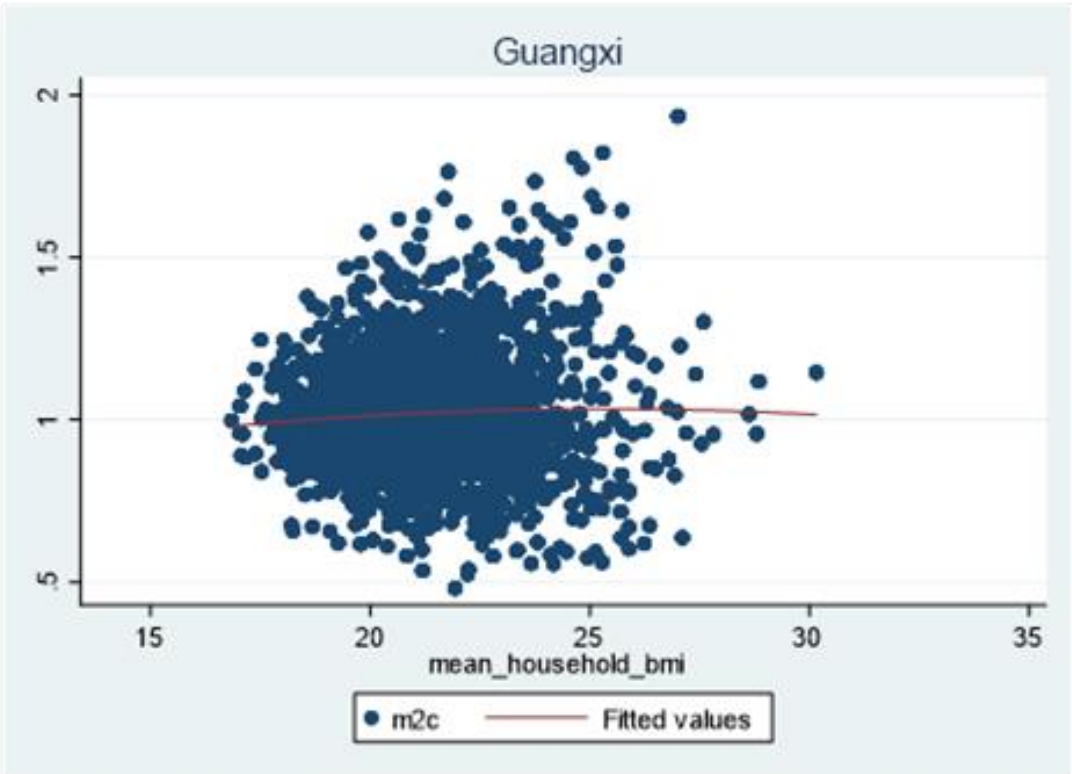
(g)


Hubei
2
1.5
1
.5
15
20
25
30
35
mean_household_bmi
m2c
Fitted values

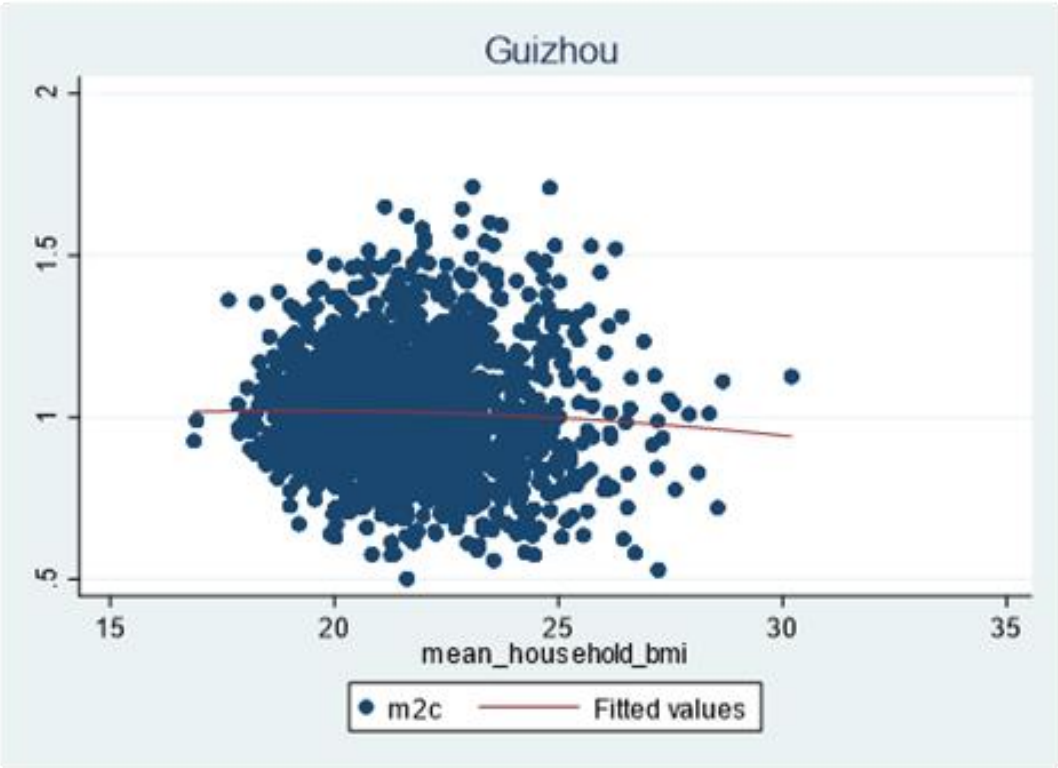
(h)


Hunan
2
1.5
1
.5
15
20
25
30
35
mean_household_bmi
m2c
Fitted v alues

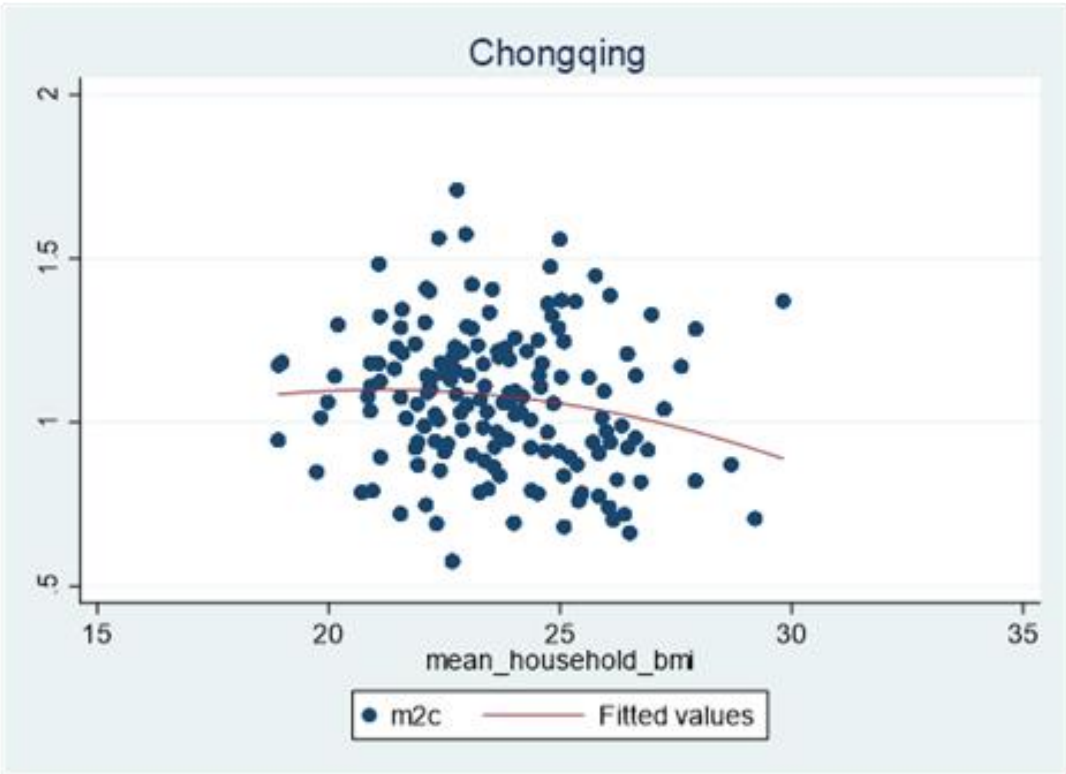
(i)


Guangxi
2
1.5
1
.5
15
20
25
30
35
mean_household_bmi
m2c
Fitted values

(j)



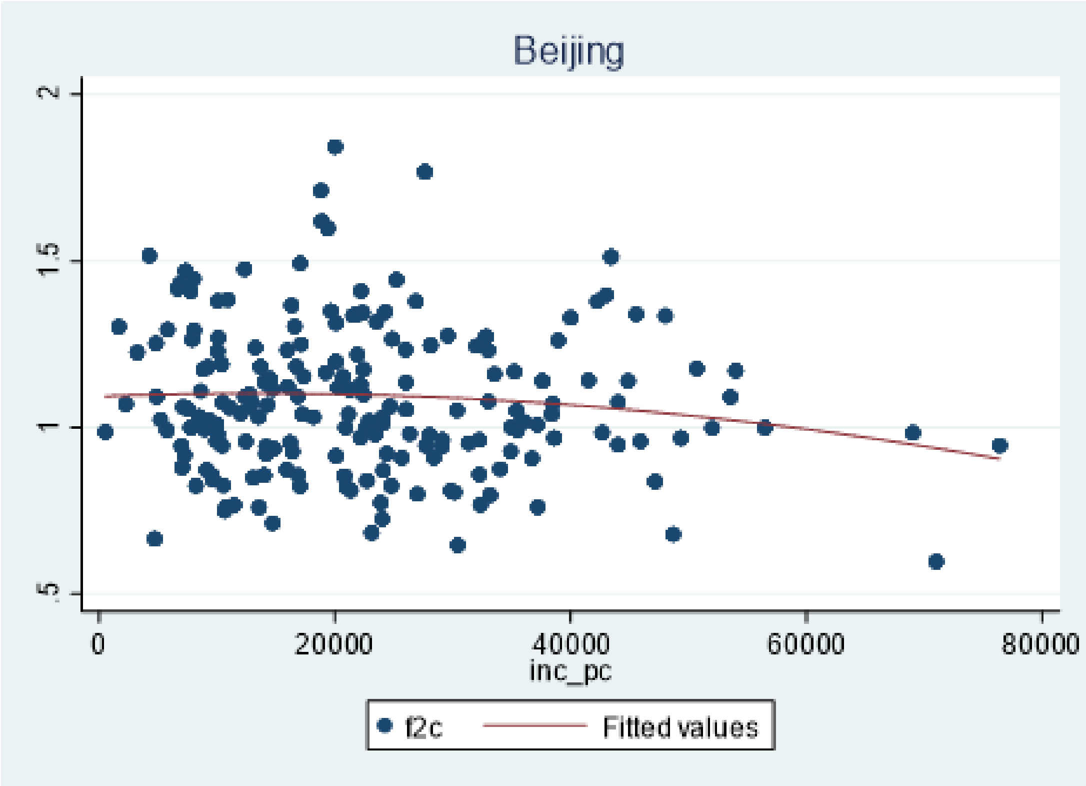
(k)



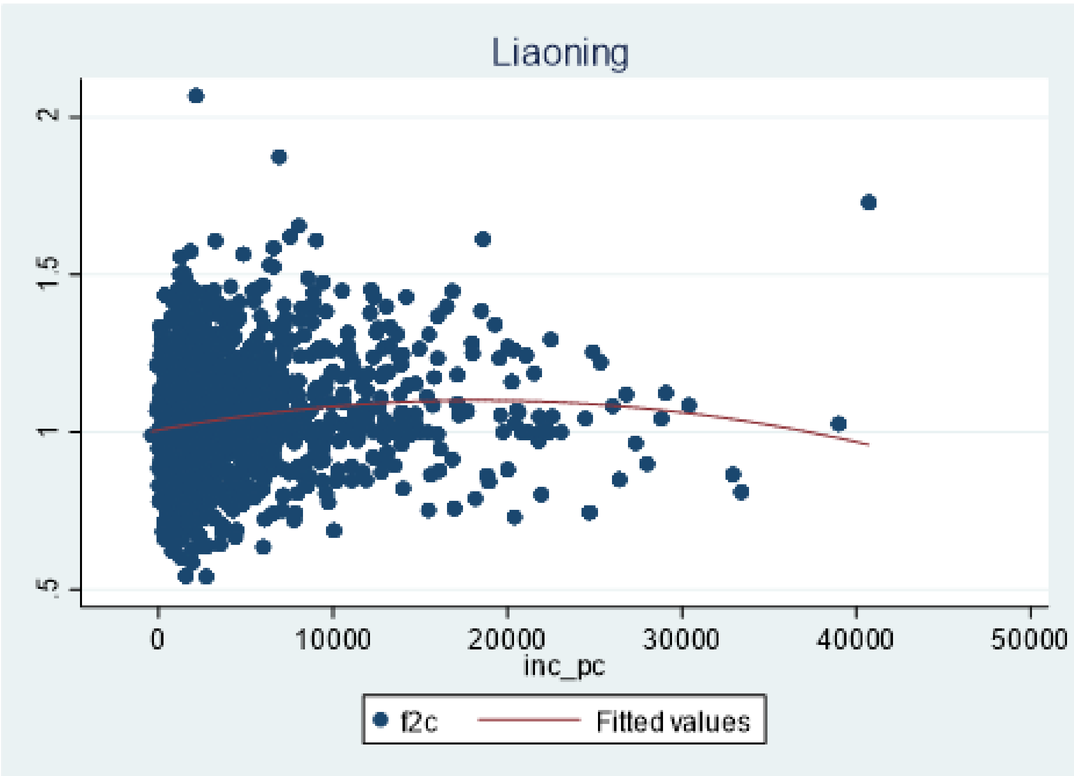
(l)

Figure III.5 (a)~(l). Mother-to-child BMI ratio ordered by mean household BMI for available years from 1898 to 2011.

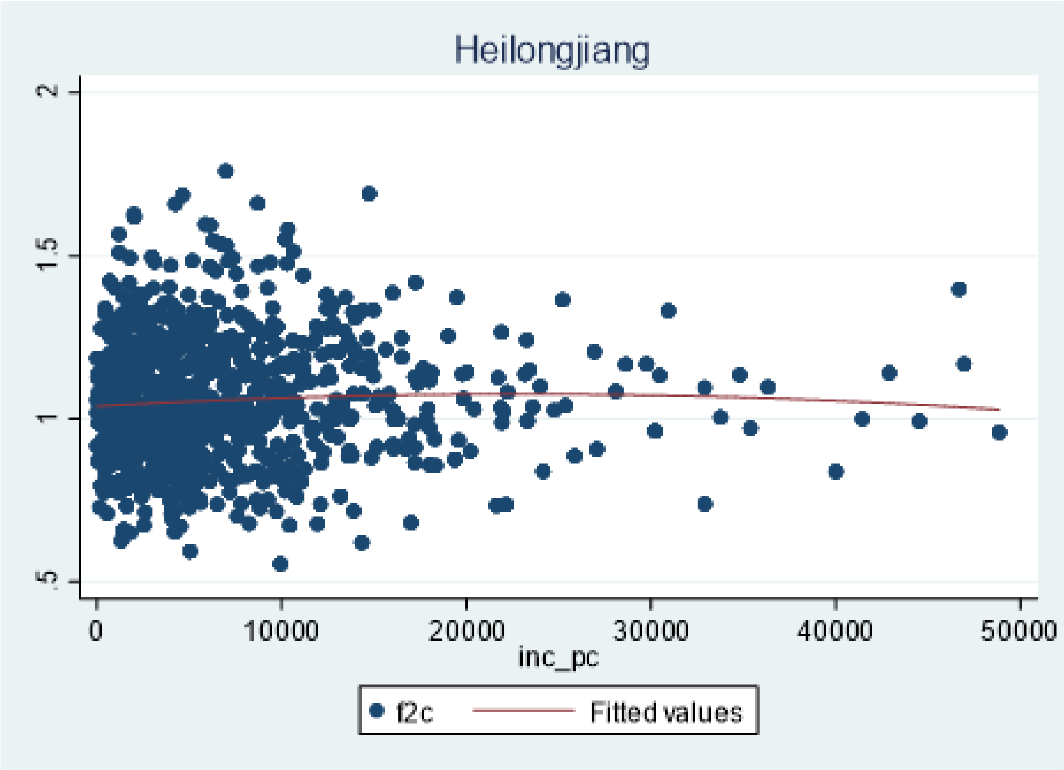
Source: CHNS1989-2011


Beijing
.5
1
1.5
2
0
20000
40000
60000
80000
inc_pc
f2c
Fitted values

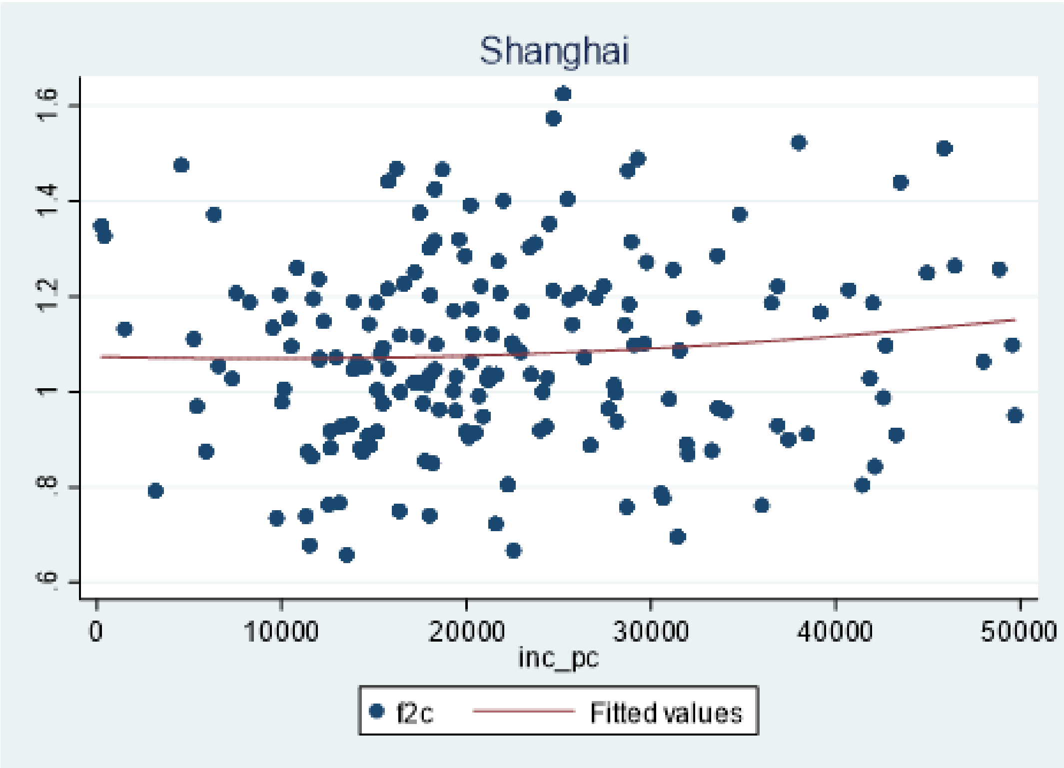
(a)


Liaoning
.5
1
1.5
2
0
10000
20000
30000
40000
50000
inc_pc
f2c
Fitted values

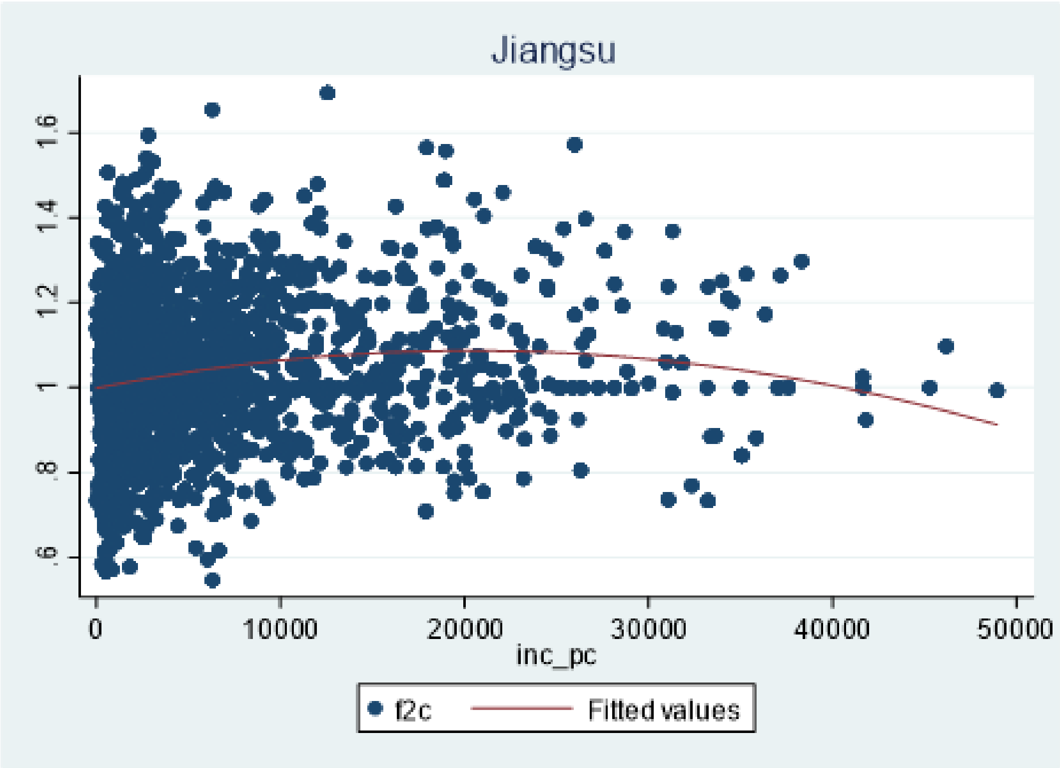
(b)


Heilongjiang
2
1.5
1
.5
0
10000
20000
30000
40000
50000
inc_pc
f2c
Fitted values

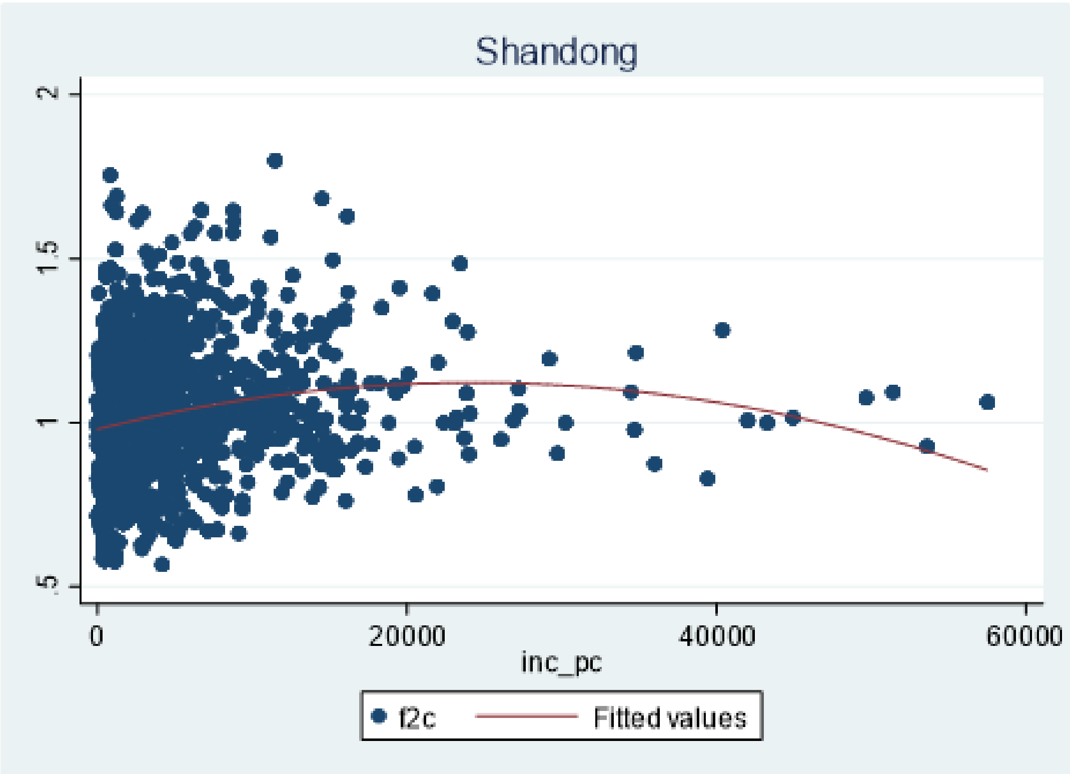
(c)


Shanghai
1.6
1.4
1.2
1
.8
.6
0
10000
20000
30000
40000
50000
inc_pc
f2c
Fitted values

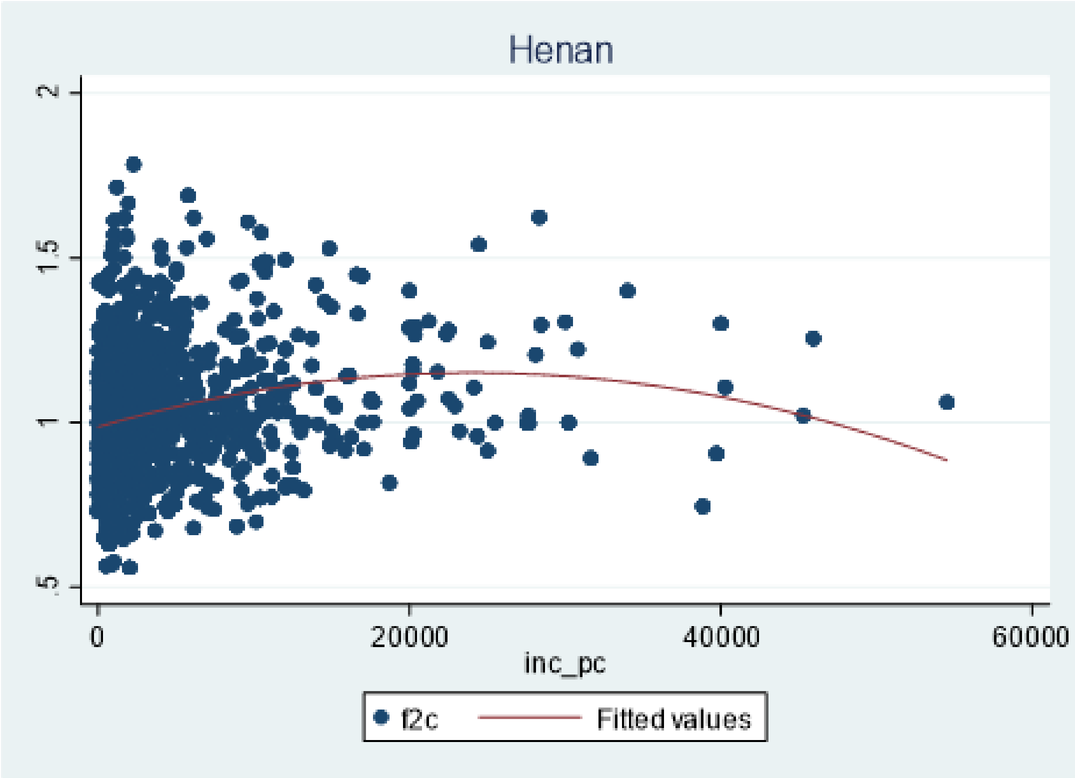
(d)


Jiangsu
1.6
1.4
1.2
1
.8
.6
0
10000
20000
30000
40000
50000
inc_pc
f2c
Fitted values

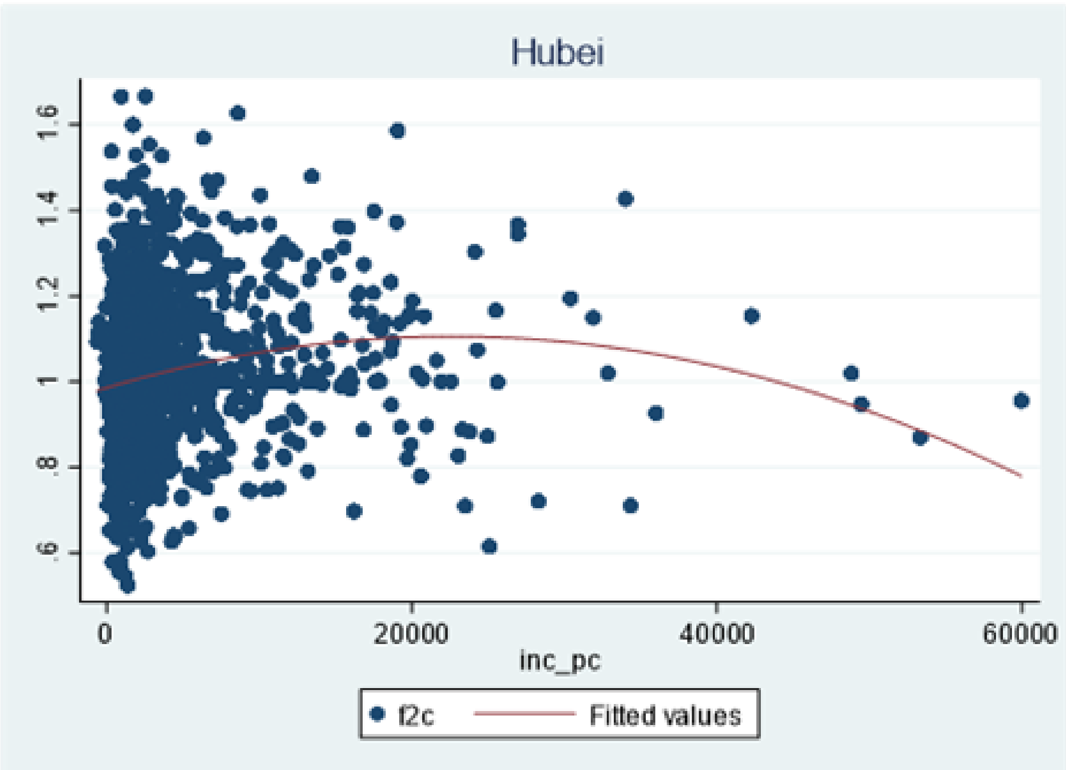
(e)


Shandong
2
1.5
1
.5
0
20000
40000
60000
inc_pc
f2c
Fitted values

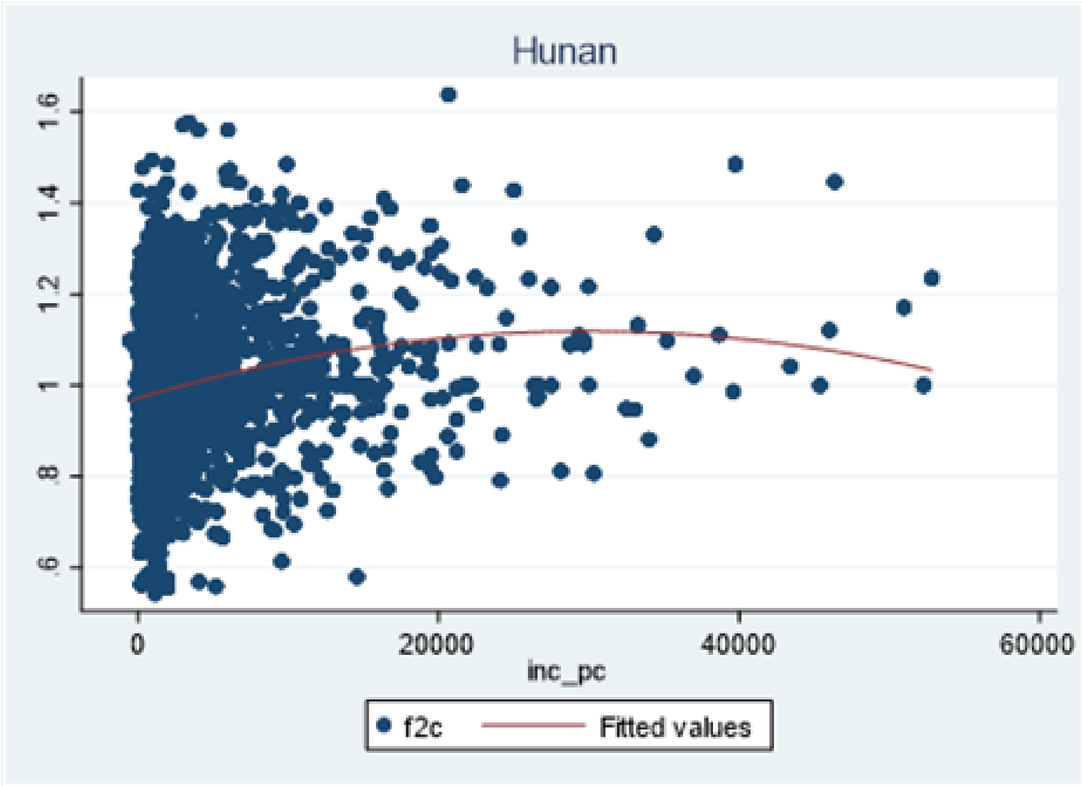
(f)


Henan
2
1.5
1
.5
0
20000
40000
60000
inc_pc
f2c
Fitted values

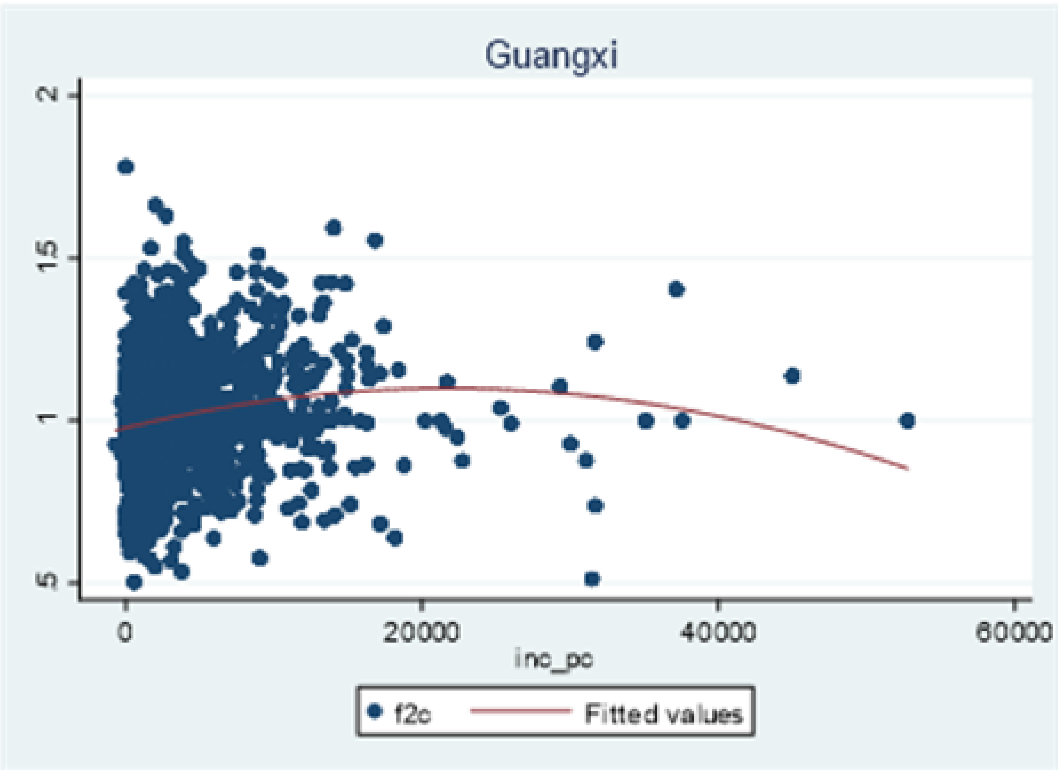
(g)


Hubei
1.6
1.4
1.2
1
.8
.6
0
20000
40000
60000
inc_pc
f2c
Fitted values

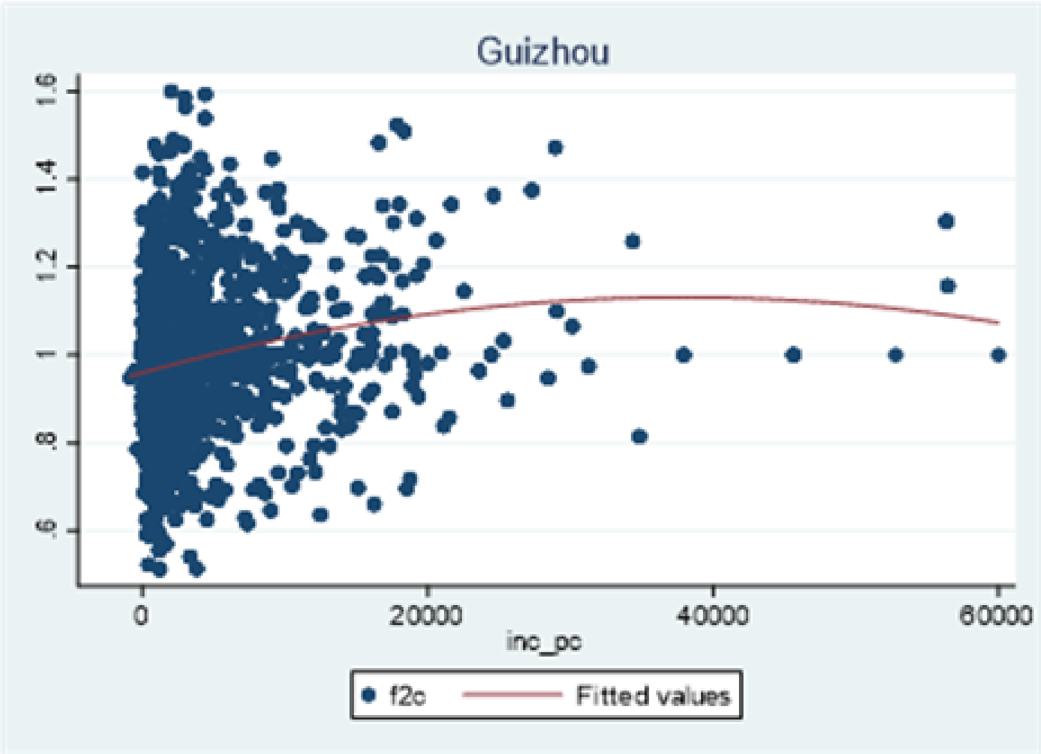
(h)


Hunan
1.6
1.4
1.2
1
.8
.6
0
20000
40000
60000
inc_pc
f2c
Fitted values

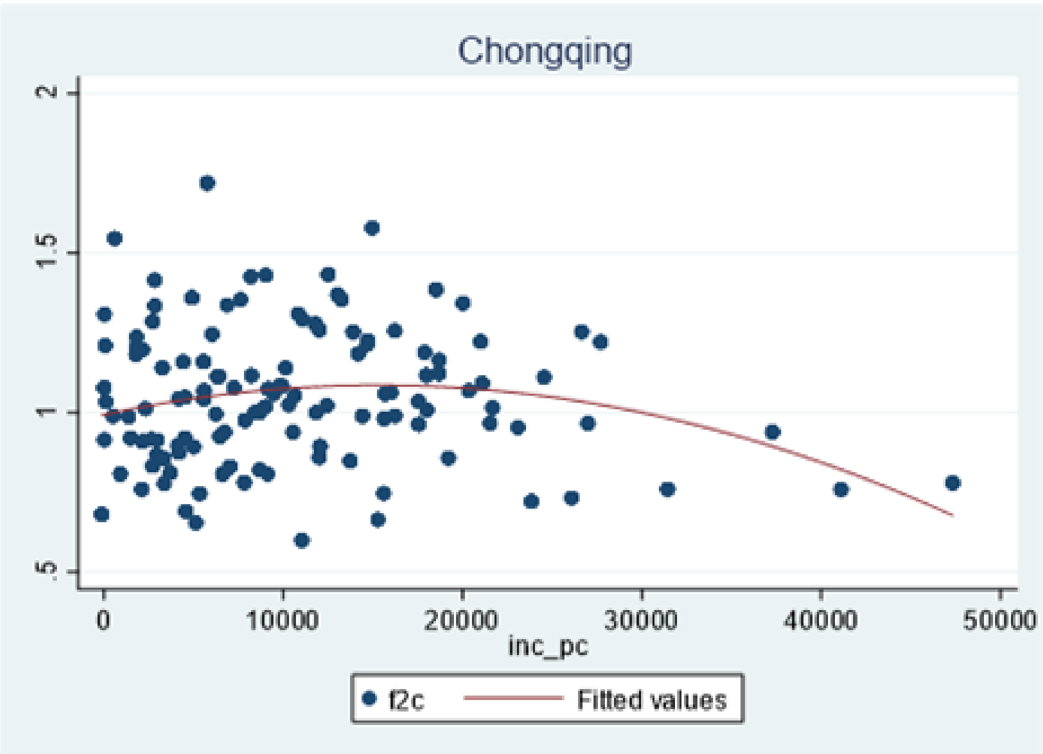
(i)


Guangxi
2
1.5
1
.5
0
20000
40000
60000
inc_pc
f2c
Fitted values

(j)

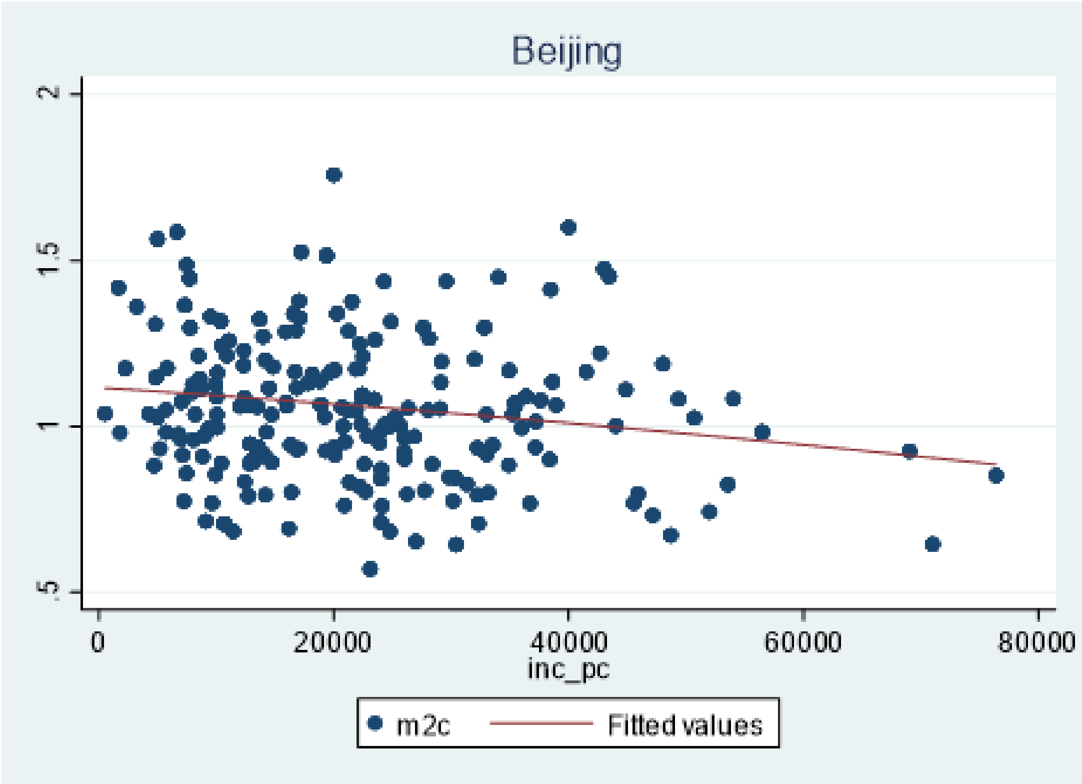


(k)



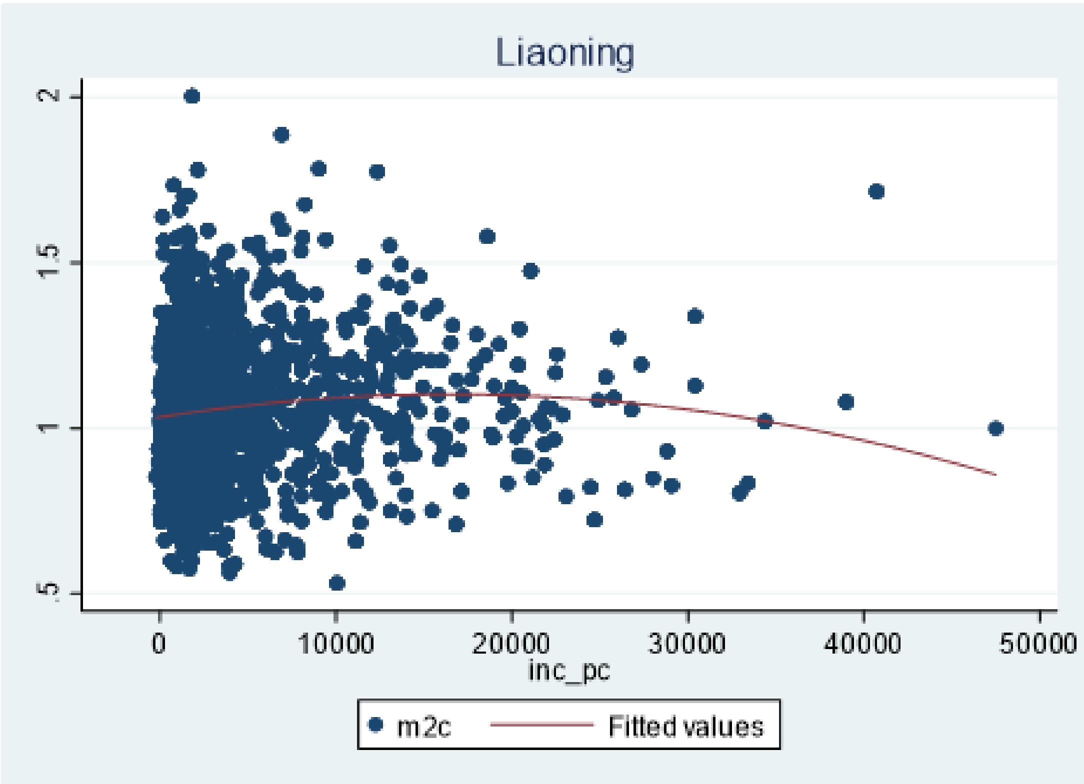
(l)

Figure III.6 (a)~(l). Father-to-child BMI ratio ordered by per capita income for available years from 1898 to 2011.
Source: CHNS1989-2011



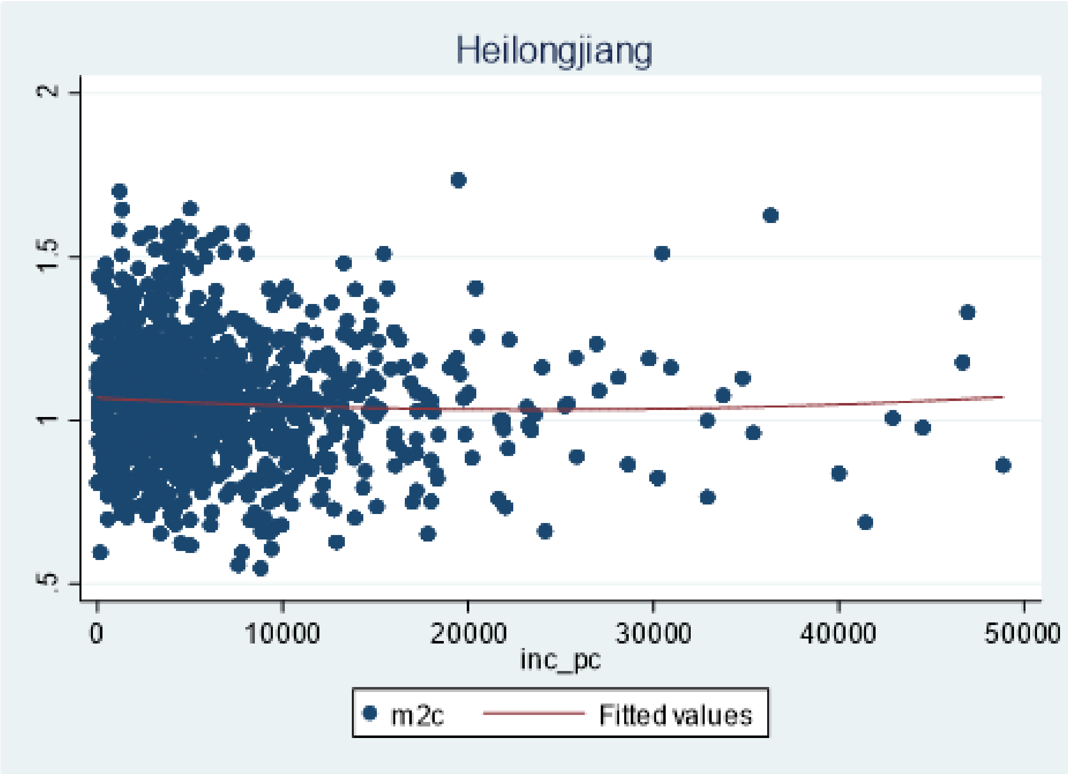
Beijing
2
1.5
1
.5
0
20000
40000
60000
80000
inc_pc
m2c
Fitted values

(a)

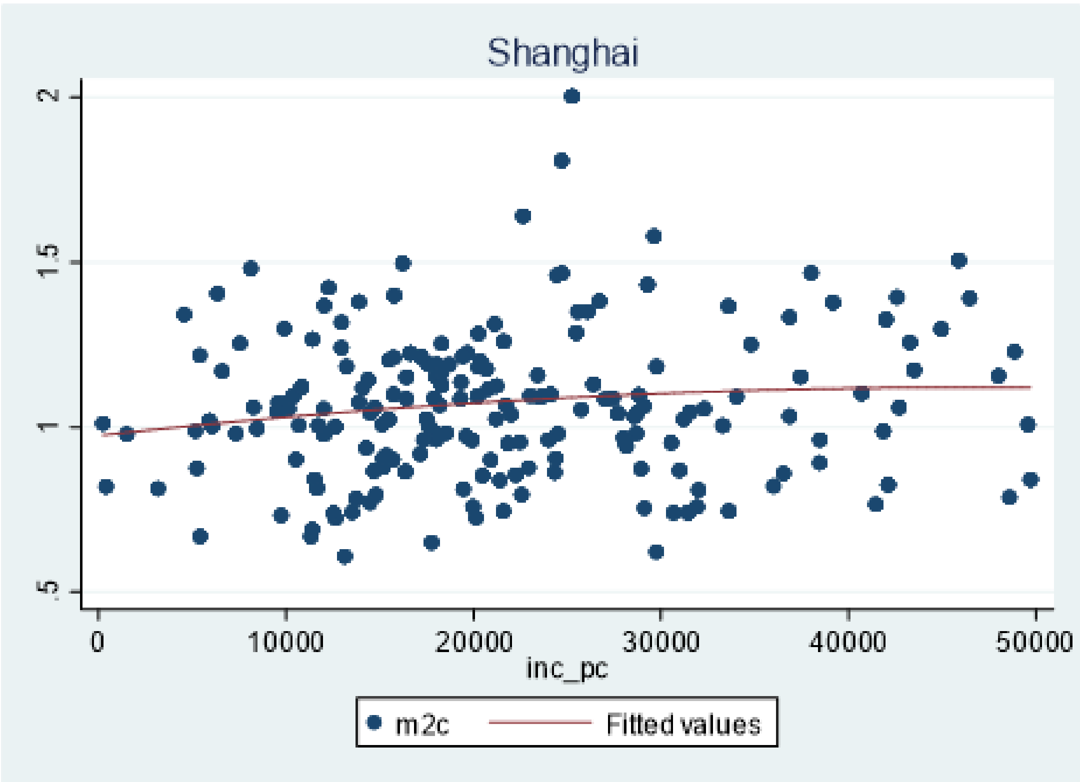


Liaoning
2
1.5
1
.5
0
10000
20000
30000
40000
50000
inc_pc
m2c
Fitted values

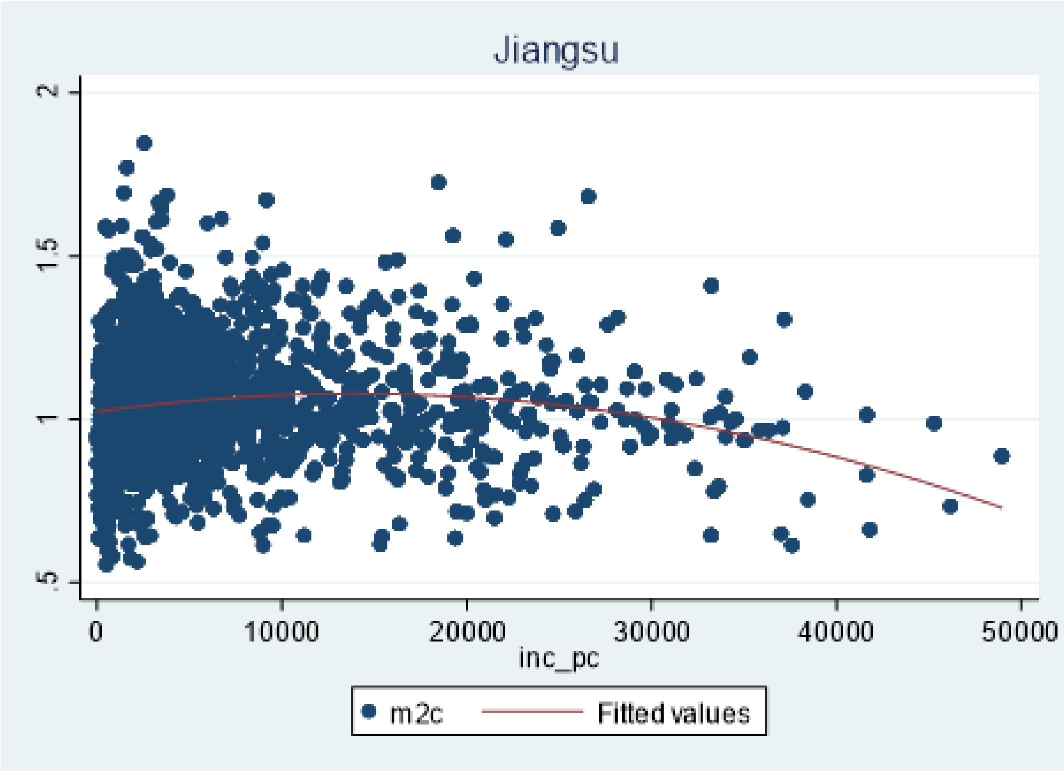
(b)


Heilongjiang
2
1.5
1
.5
0
10000
20000
30000
40000
50000
inc_pc
m2c
Fitted values

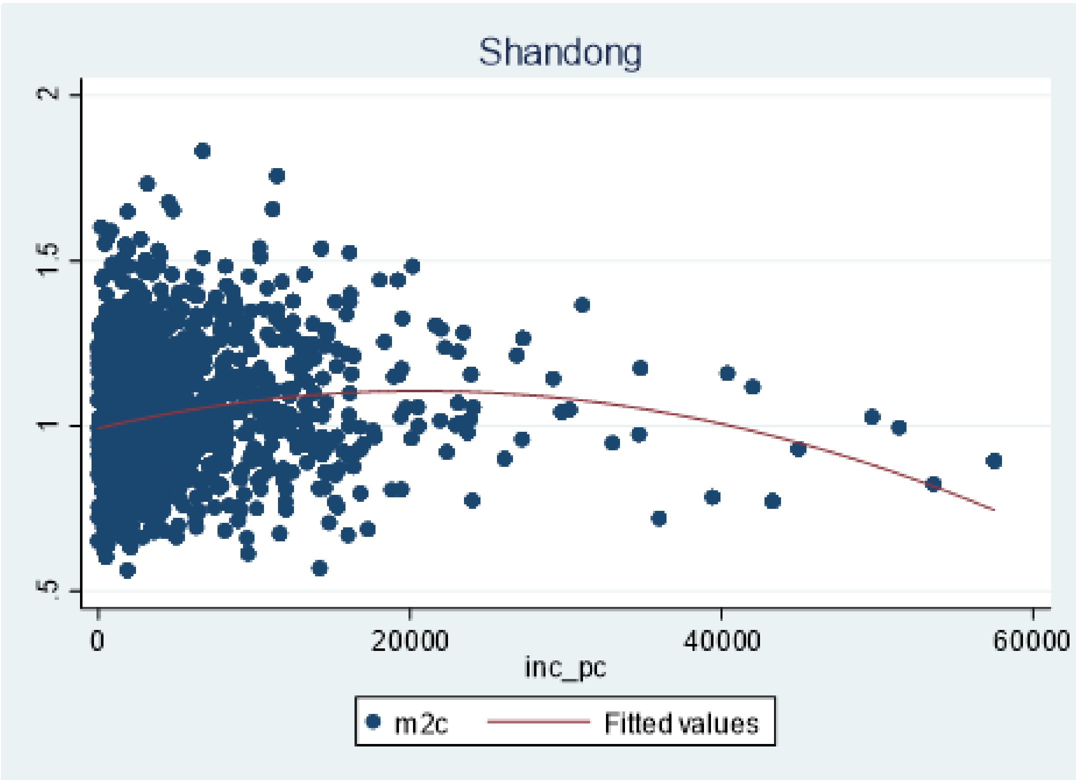
(c)


Shanghai
2
1.5
1
.5
0
10000
20000
30000
40000
50000
inc_pc
m2c
Fitted values

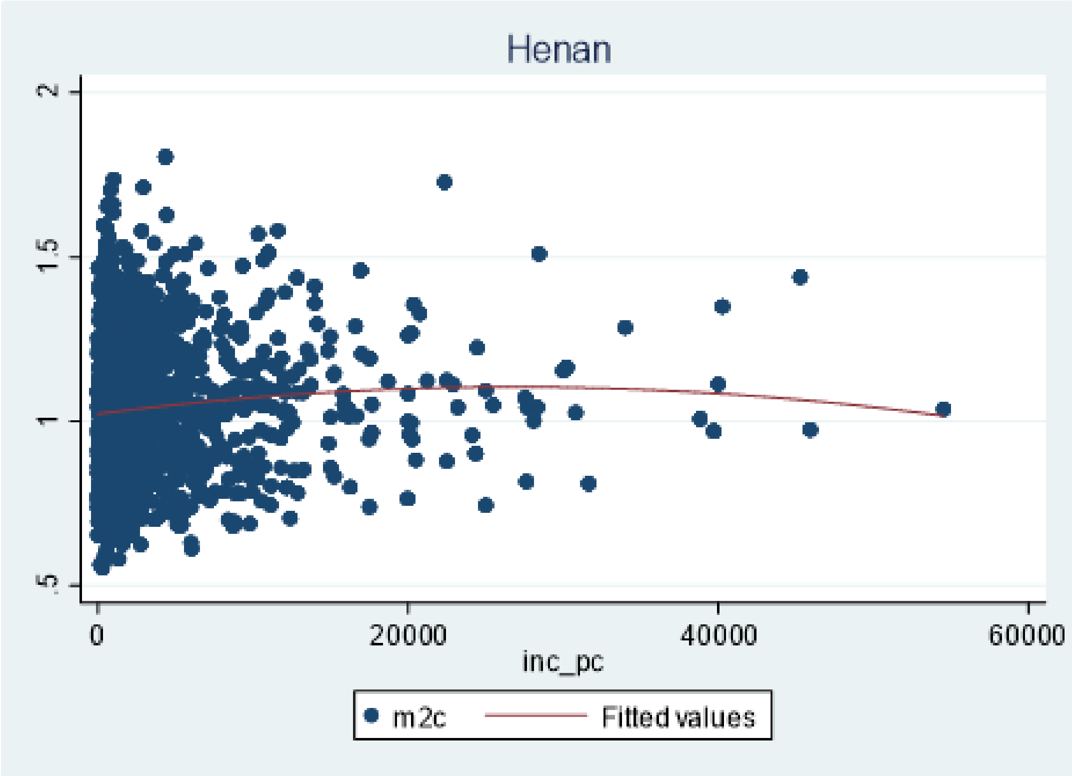
(d)


Jiangsu
2
1.5
1
.5
0
10000
20000
30000
40000
50000
inc_pc
m2c
Fitted values

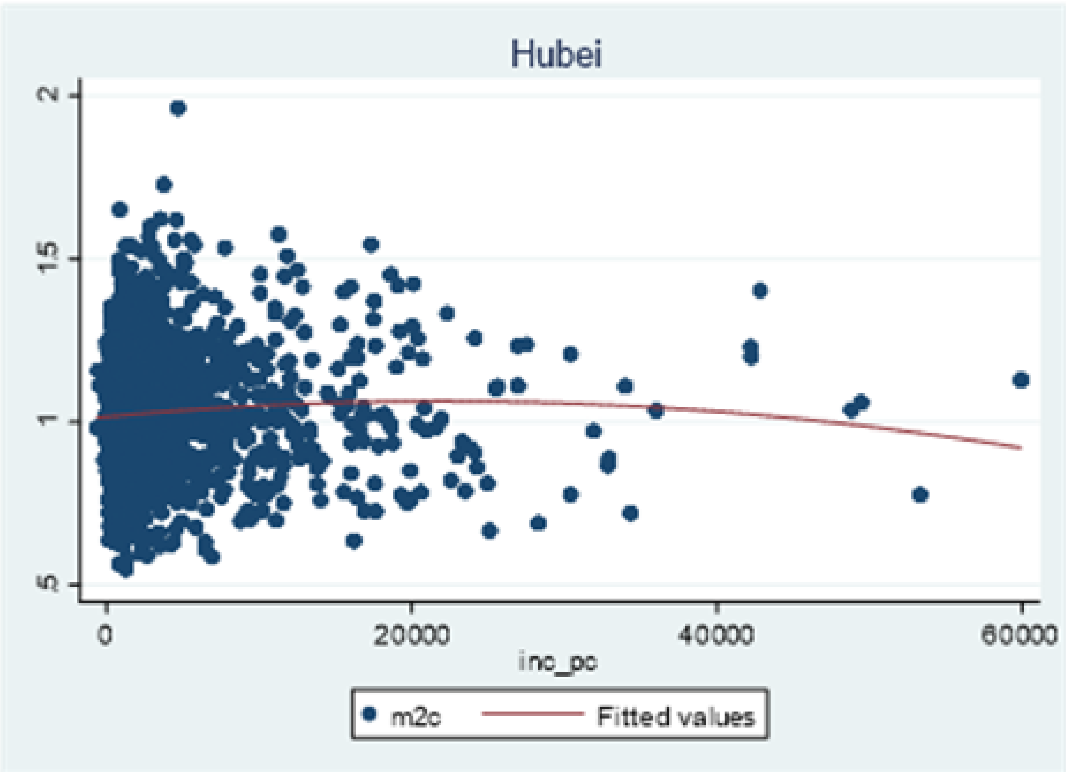
(e)


Shandong
2
1.5
1
.5
0
20000
40000
60000
inc_pc
m2c
Fitted values

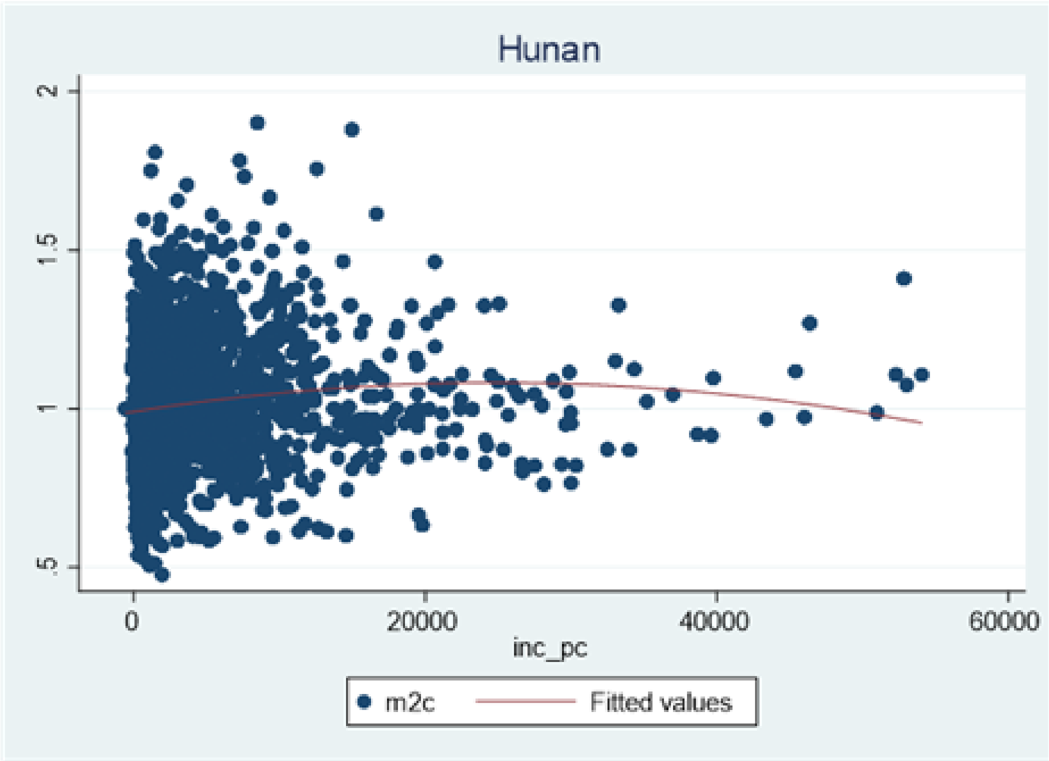
(f)


Henan
2
1.5
1
.5
0
20000
40000
60000
inc_pc
m2c
Fitted values

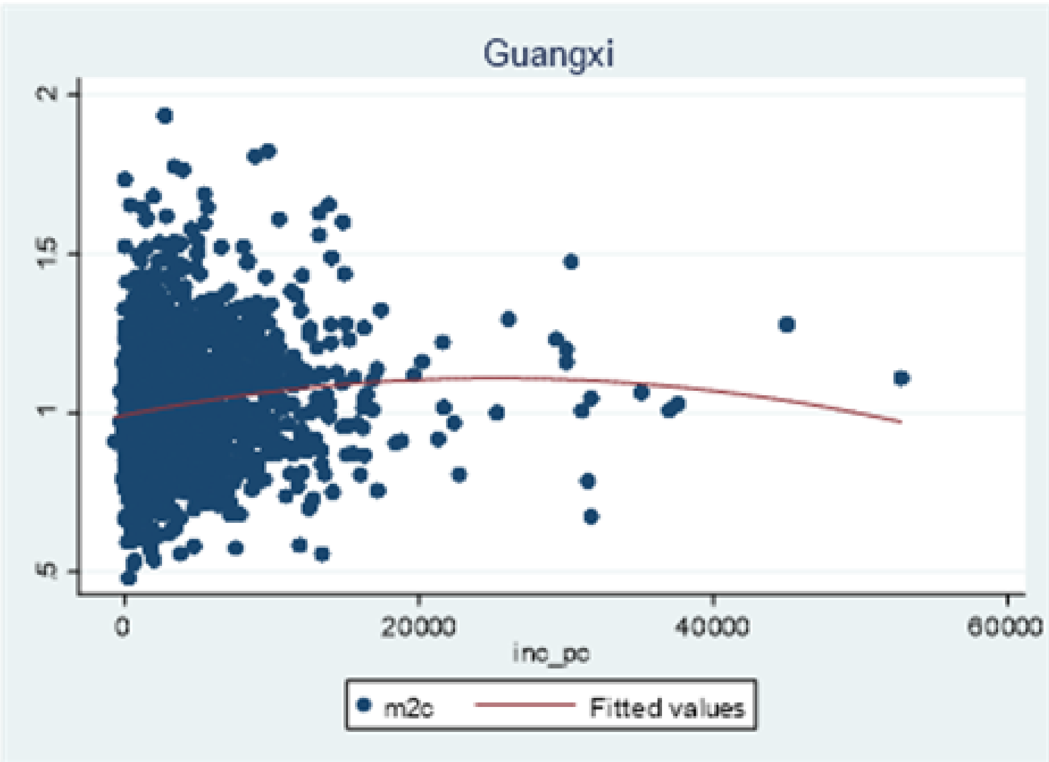
(g)


Hubei
2
1.5
1
.5
0
20000
40000
60000
inc_pc
m2c
Fitted values

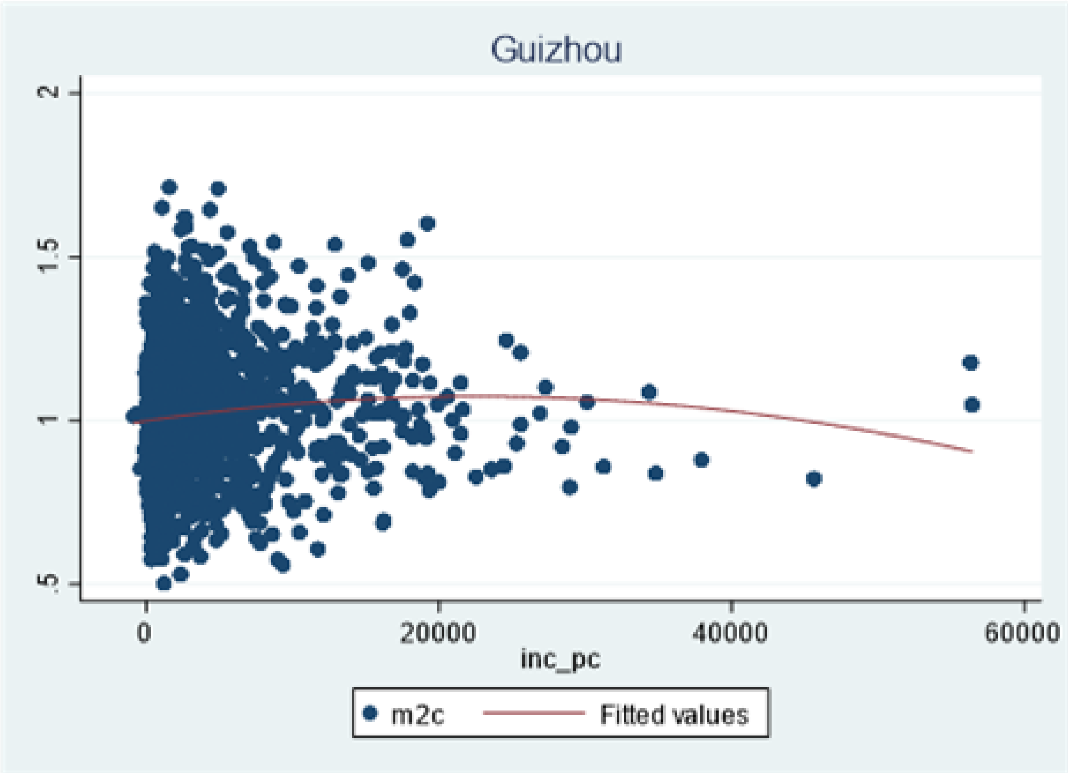
(h)


Hunan
.5
1
1.5
2
0
20000
40000
60000
inc_pc
m2c
Fitted values

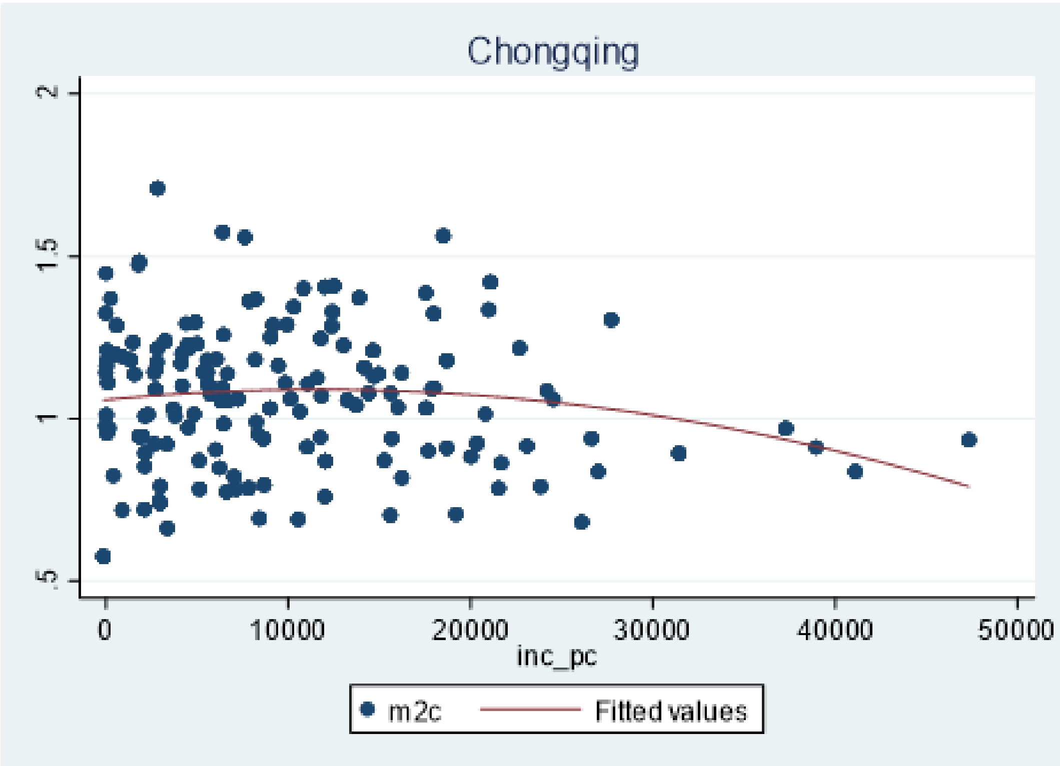
(i)


Guangxi
.5
1
1.5
2
0
20000
40000
60000
inc_pc
m2c
Fitted values

(j)



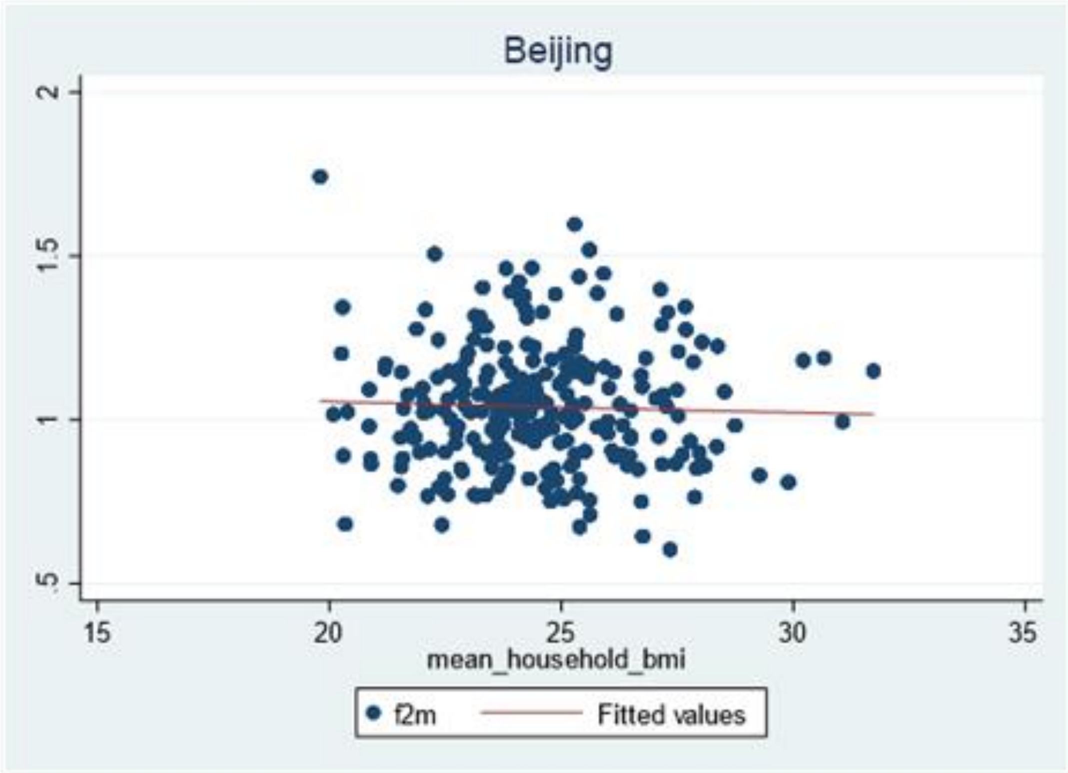
(k)



(l)

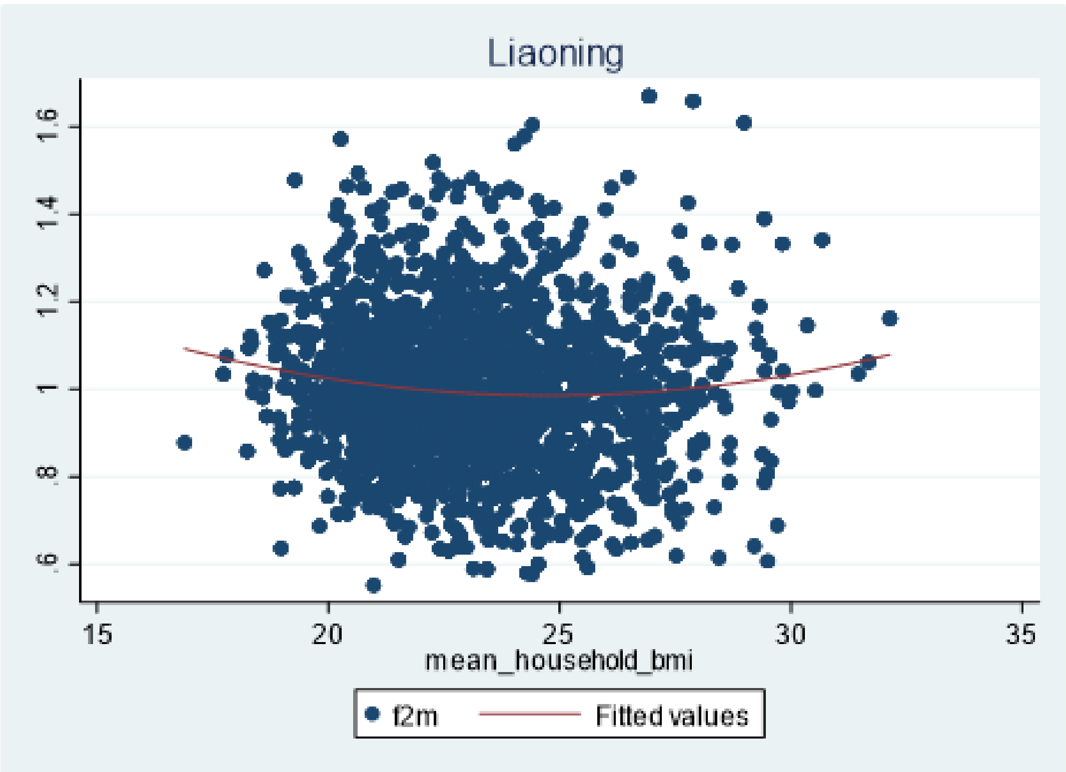
Figure III.7 (a)~(l). Mother to child BMI ratio ordered by per capita income for available years from 1898 to 2011.

Source: CHNS1989-2011



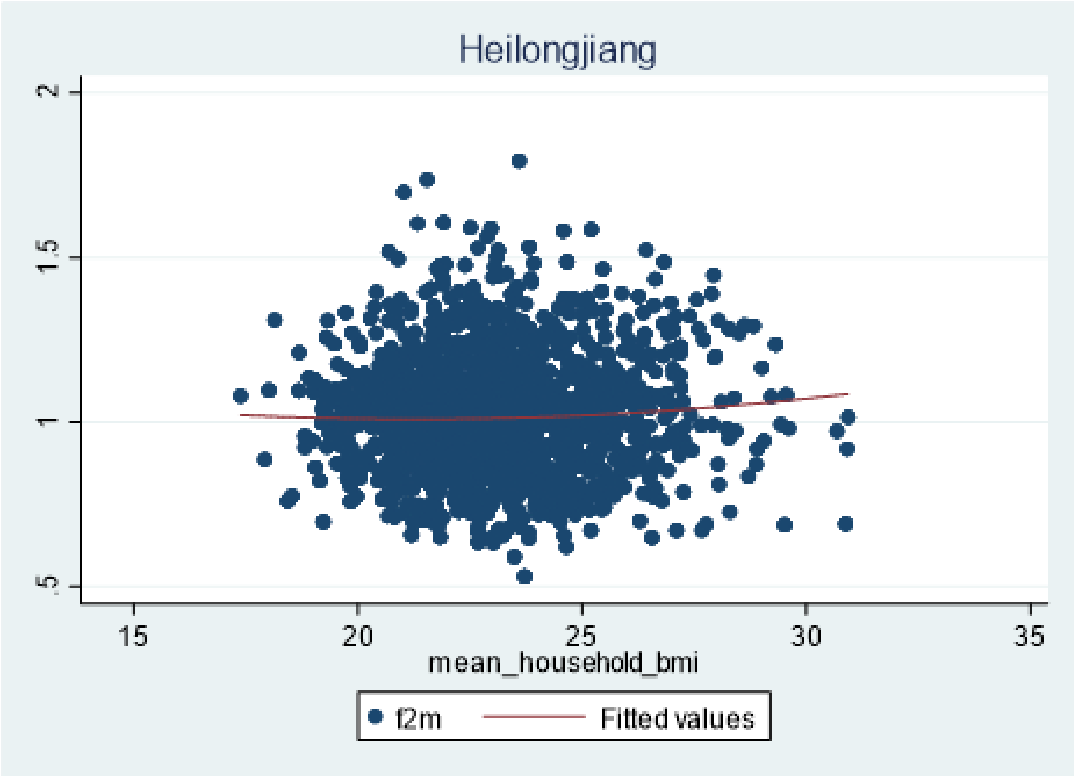
Beijing
2
1.5
1
.5
15
20
25
30
35
mean_household_bmi
f2m
Fitted values

(a)

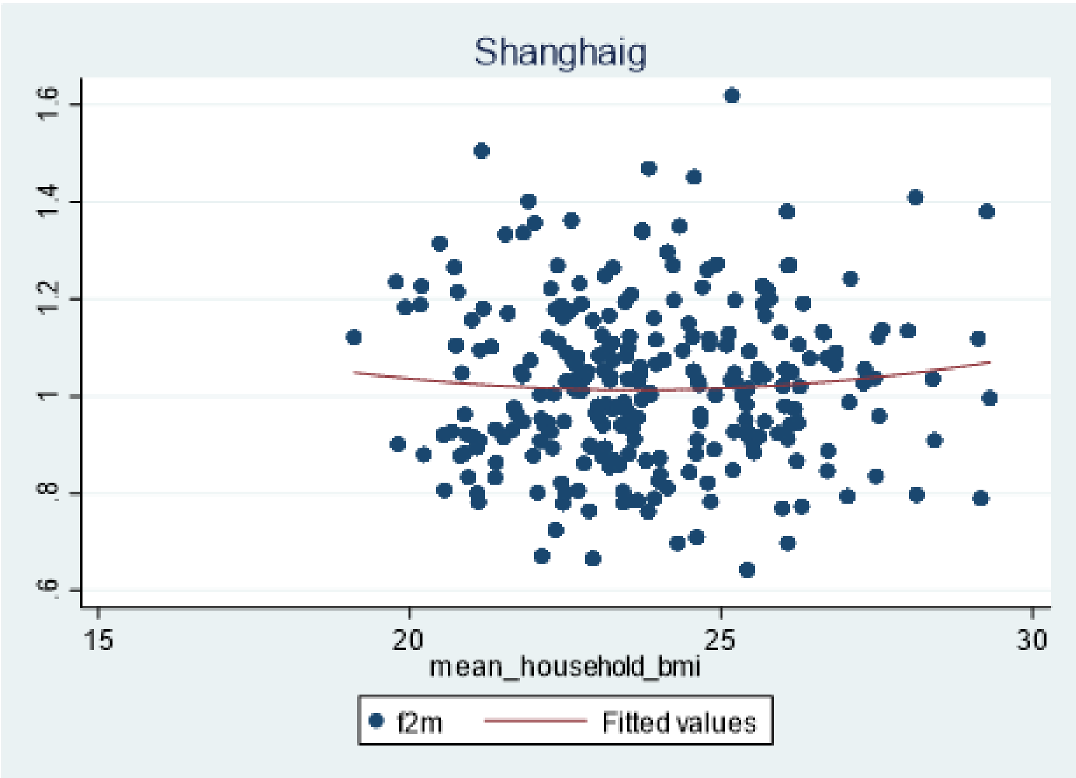


Liaoning
1.6
1.4
1.2
1
.8
.6
15
20
25
30
35
mean_household_bmi
f2m
Fitted values

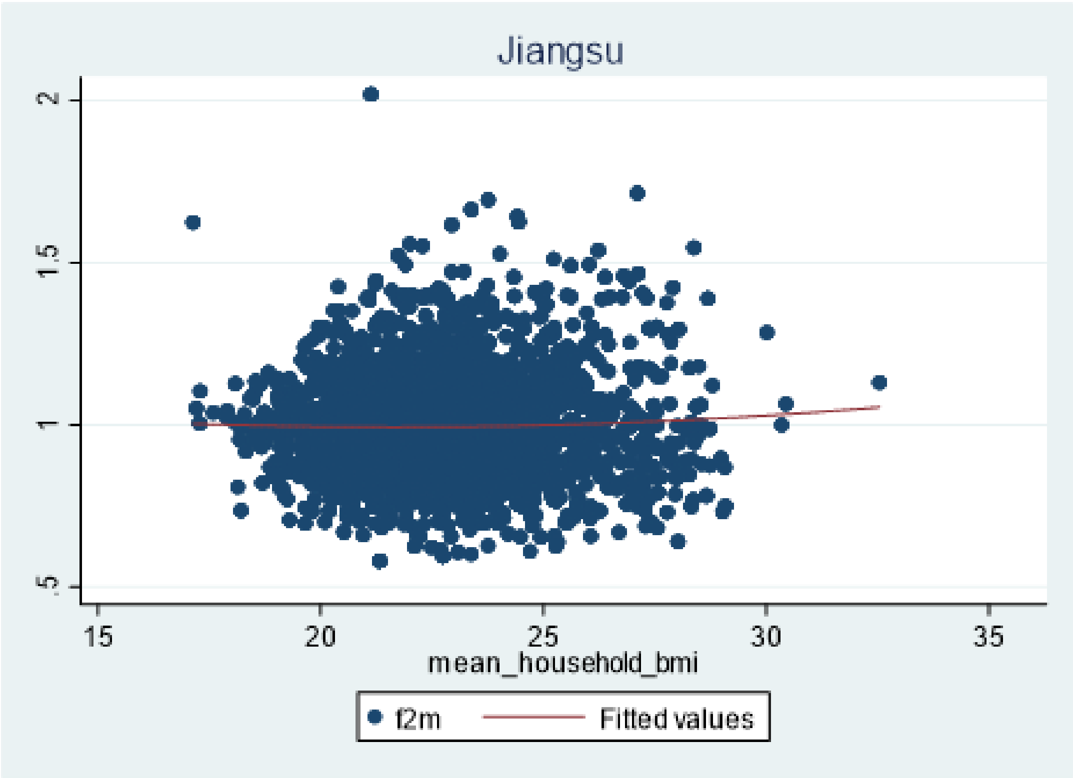
(b)


Heilongjiang
2
1.5
1
.5
15
20
25
30
35
mean_household_bmi
f2m
Fitted values

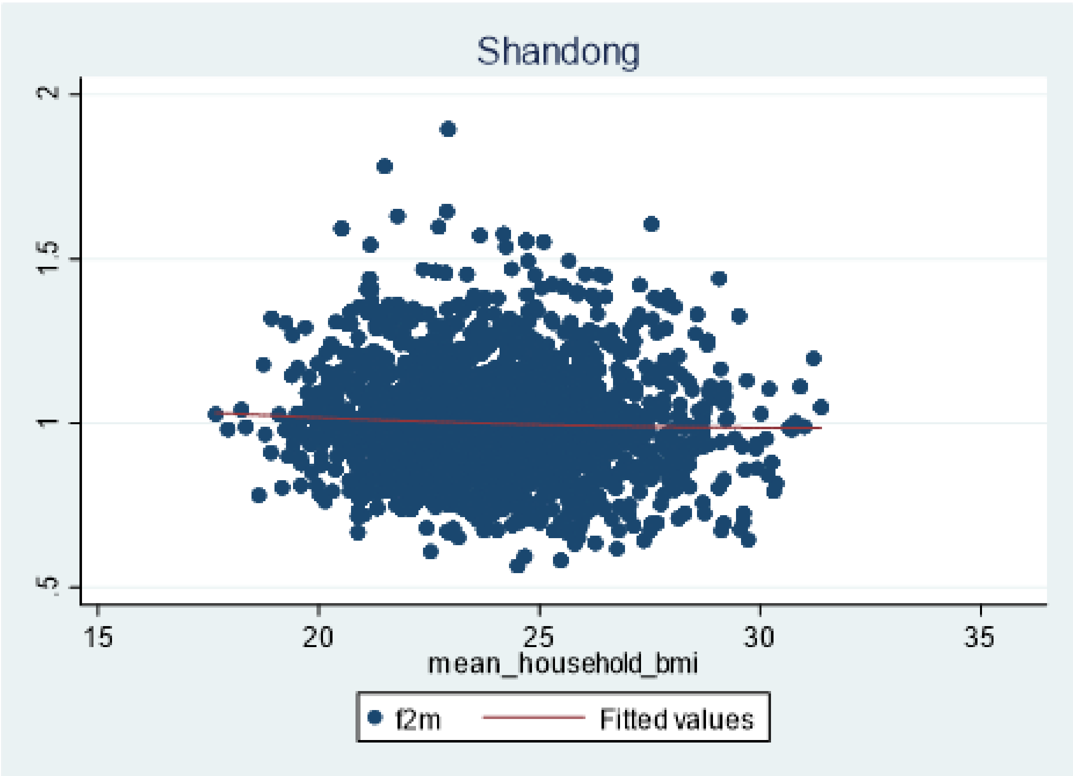
(c)


Shanghaig
1.6
1.4
1.2
1
.8
.6
15
20
25
30
mean_household_bmi
f2m
Fitted values

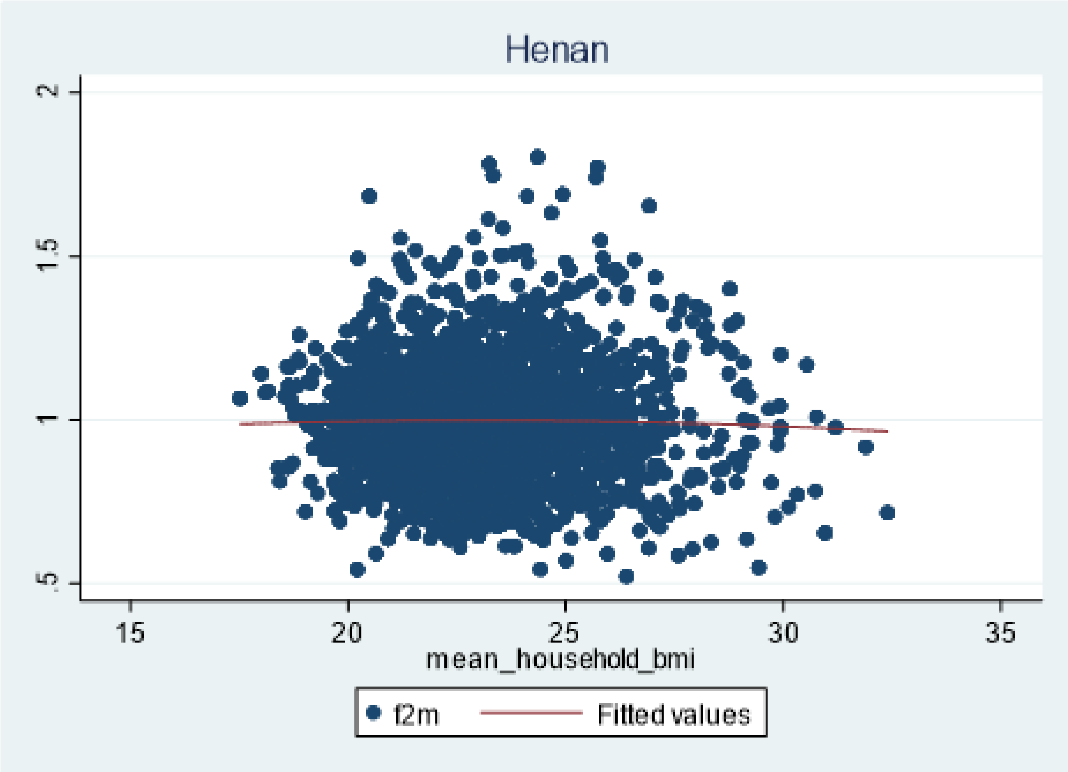
(d)


Jiangsu
2
1.5
1
.5
15
20
25
30
35
mean_household_bmi
f2m
Fitted values

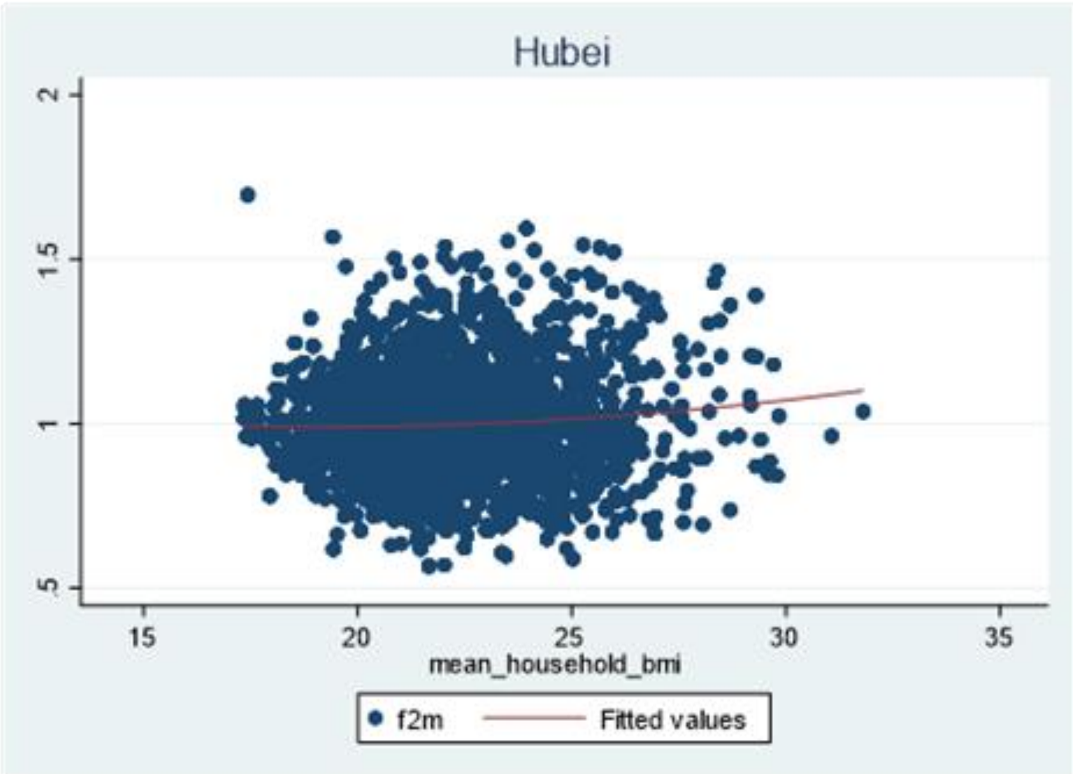
(e)


Shandong
2
1.5
1
.5
15
20
25
30
35
mean_household_bmi
f2m
Fitted values

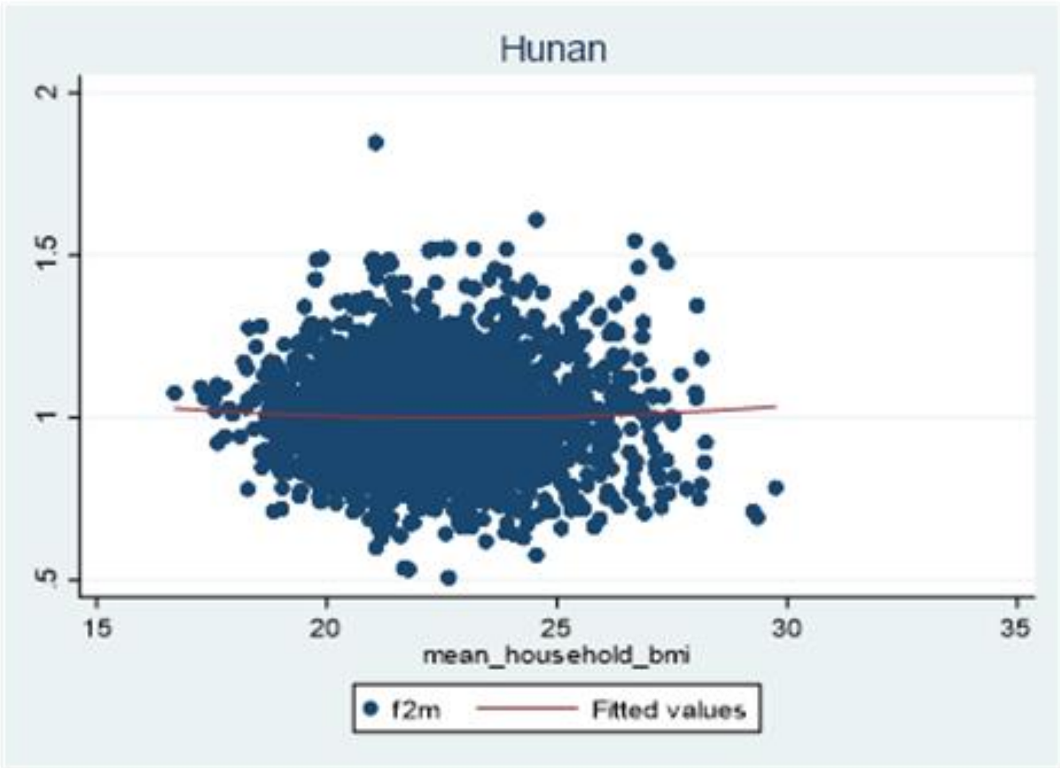
(f)


Henan
2
1.5
1
.5
15
20
25
30
35
mean_household_bmi
f2m
Fitted values

(g)

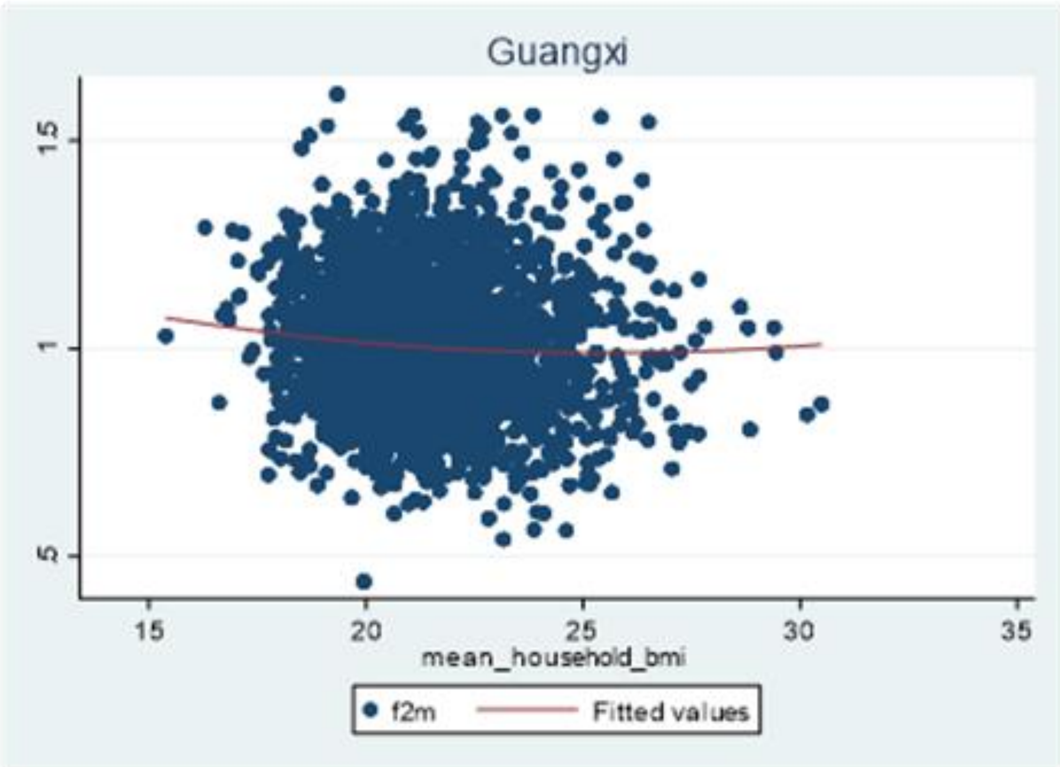

Hubei
2
1.5
1
.5
15
20
25
30
35
mean_household_bmi
f2m
Fitted values

(h)



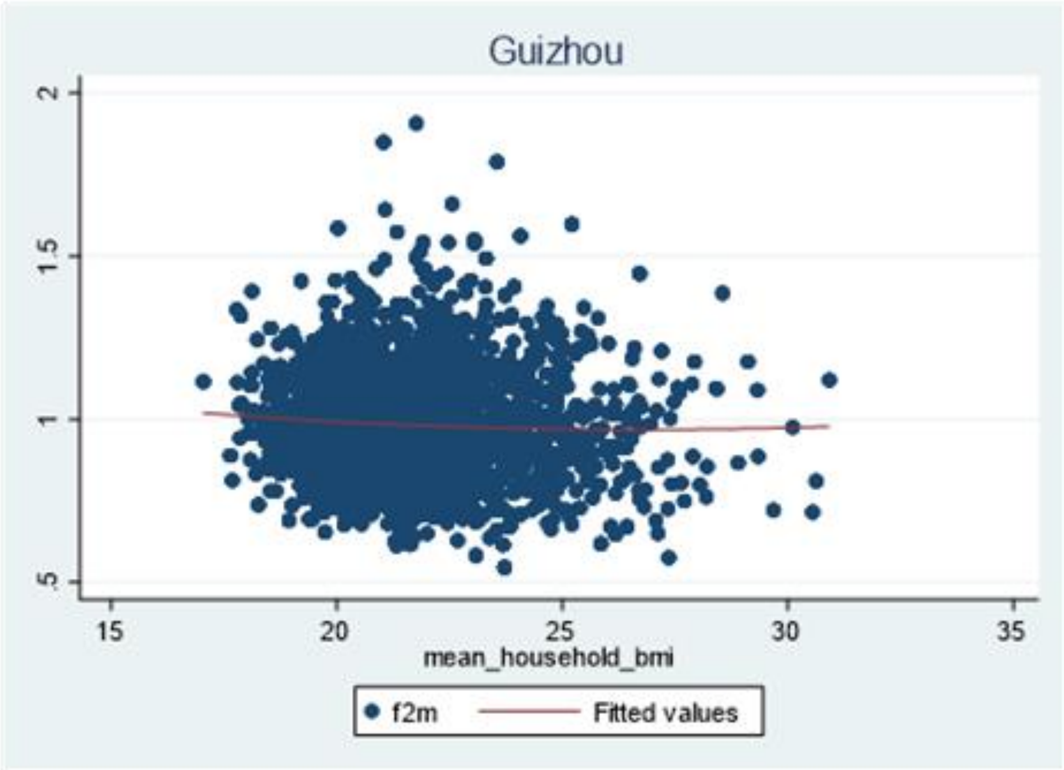
Hunan
2
1.5
1
.5
15
20
25
30
35
mean_household_bmi
f2m
Fitted values

(i)

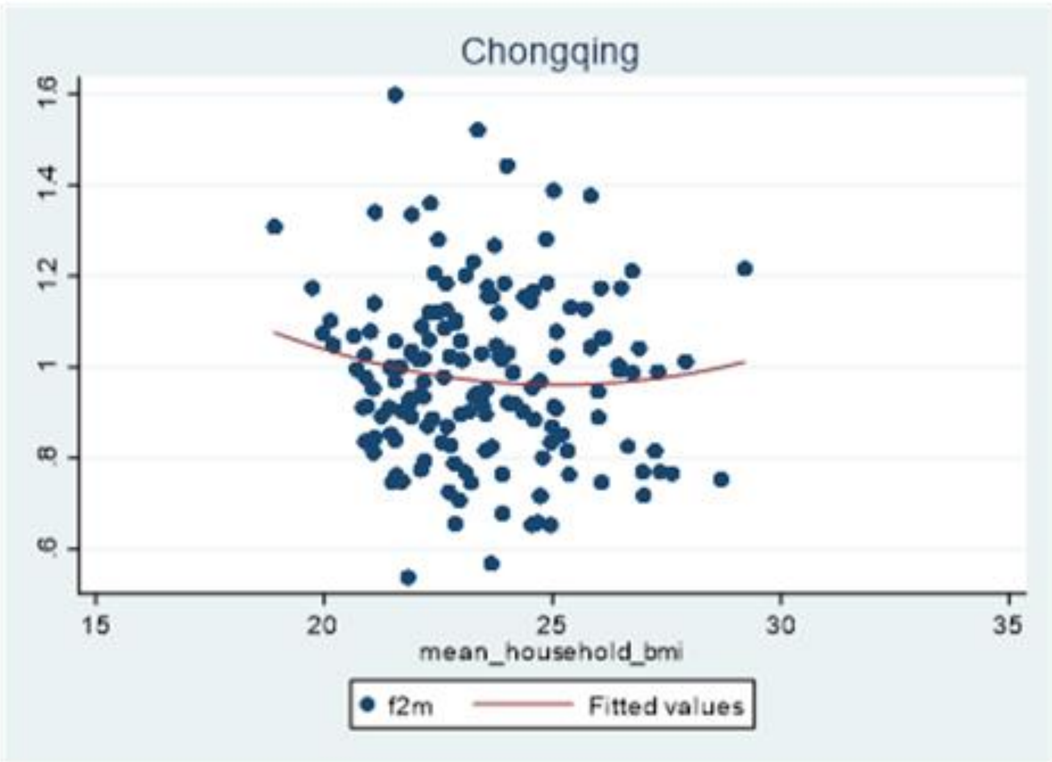


Guangxi
1.5
1
.5
15
20
25
30
35
mean_household_bmi
f2m
Fitted values

(j)



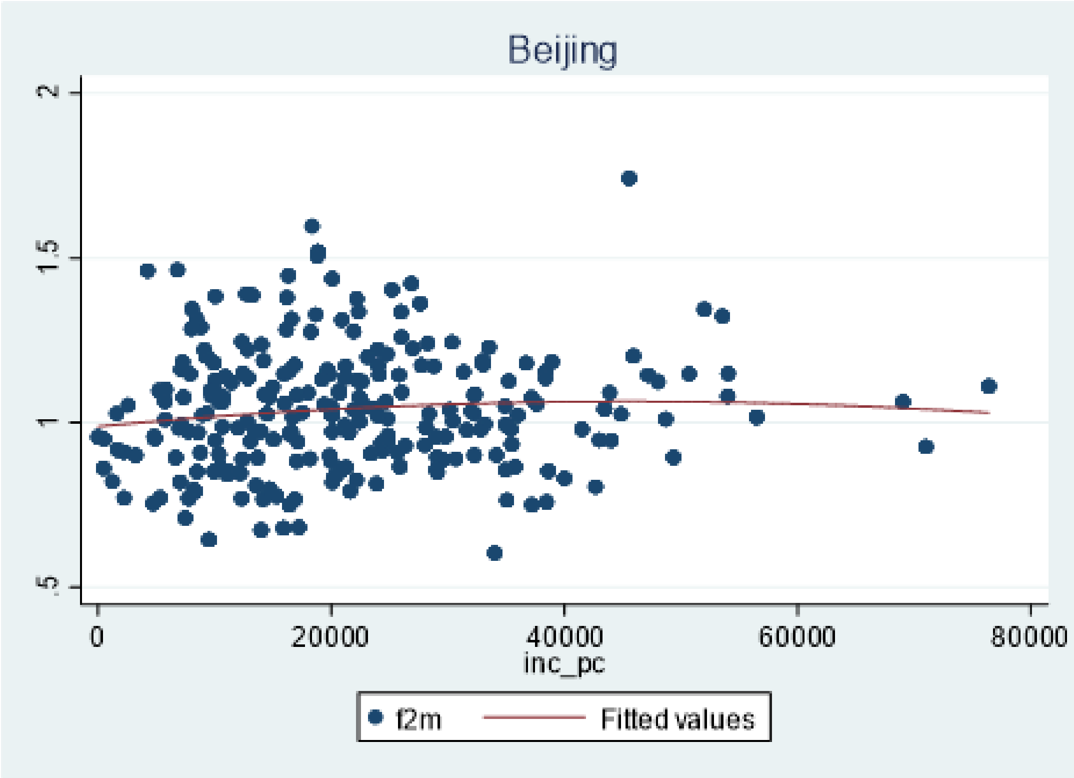
(k)



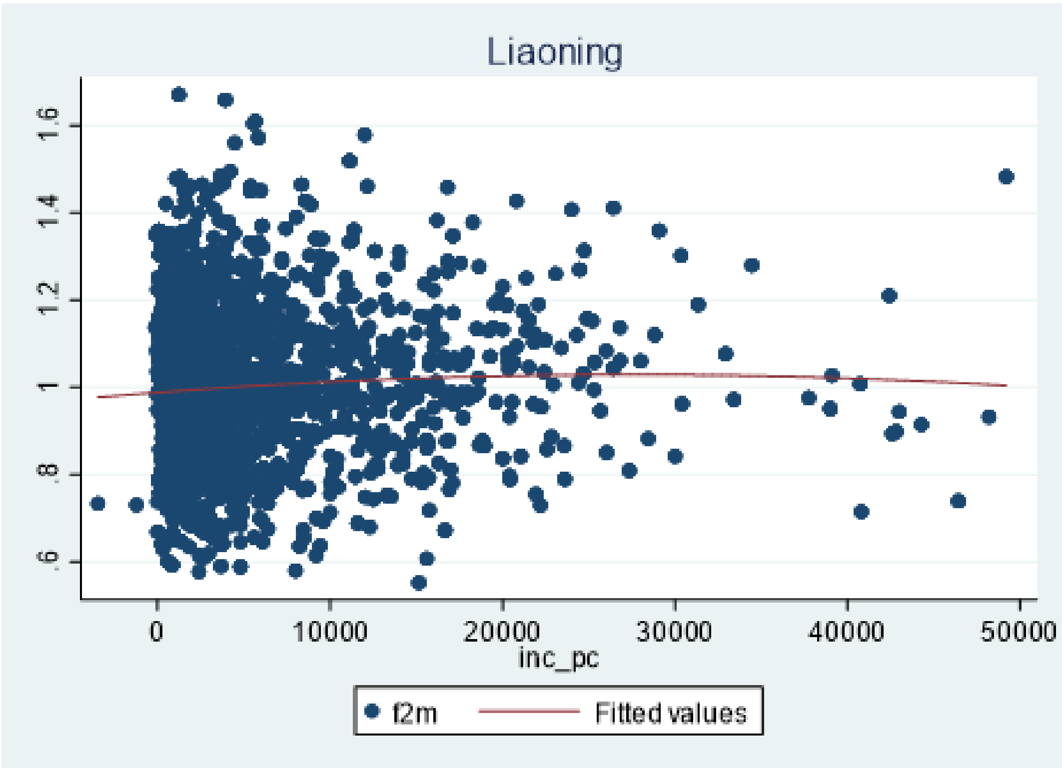
(l)

Figure III.8 (a)~(l). Male head to spouse BMI ratio ordered by mean household BMI for available years from 1898 to 2011.

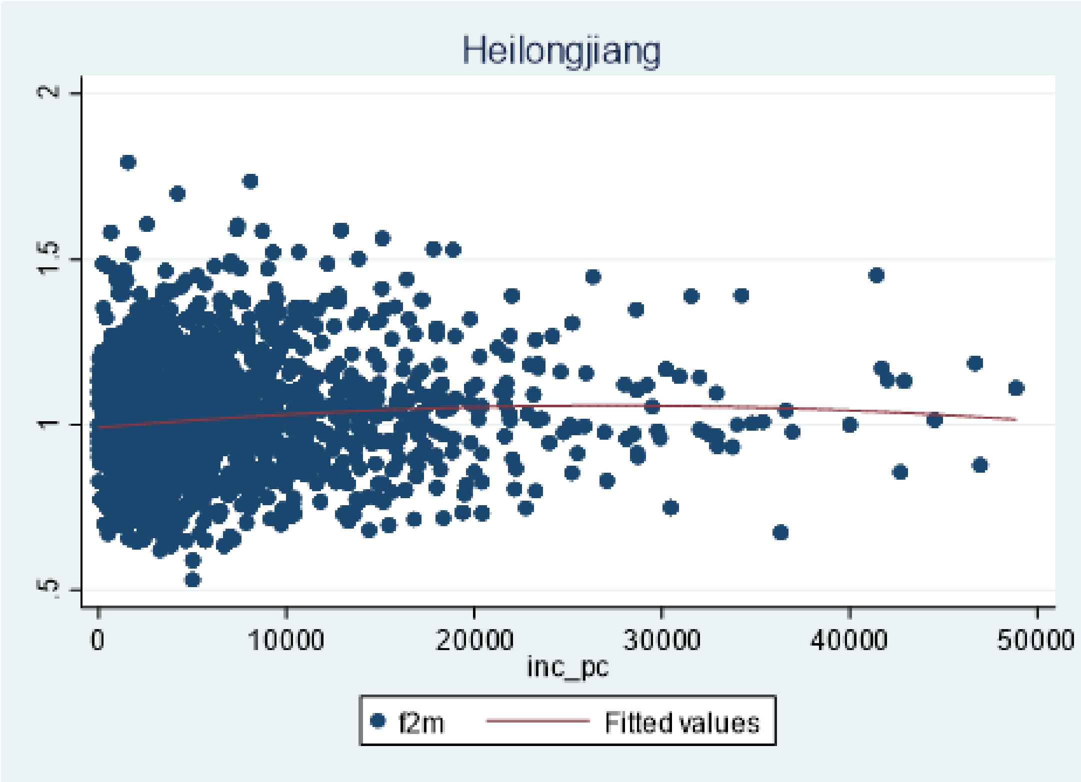
Source: CHNS1989-2011


Beijing
2
1.5
1
.5
0
20000
40000
60000
80000
inc_pc
f2m
Fitted values

(a)

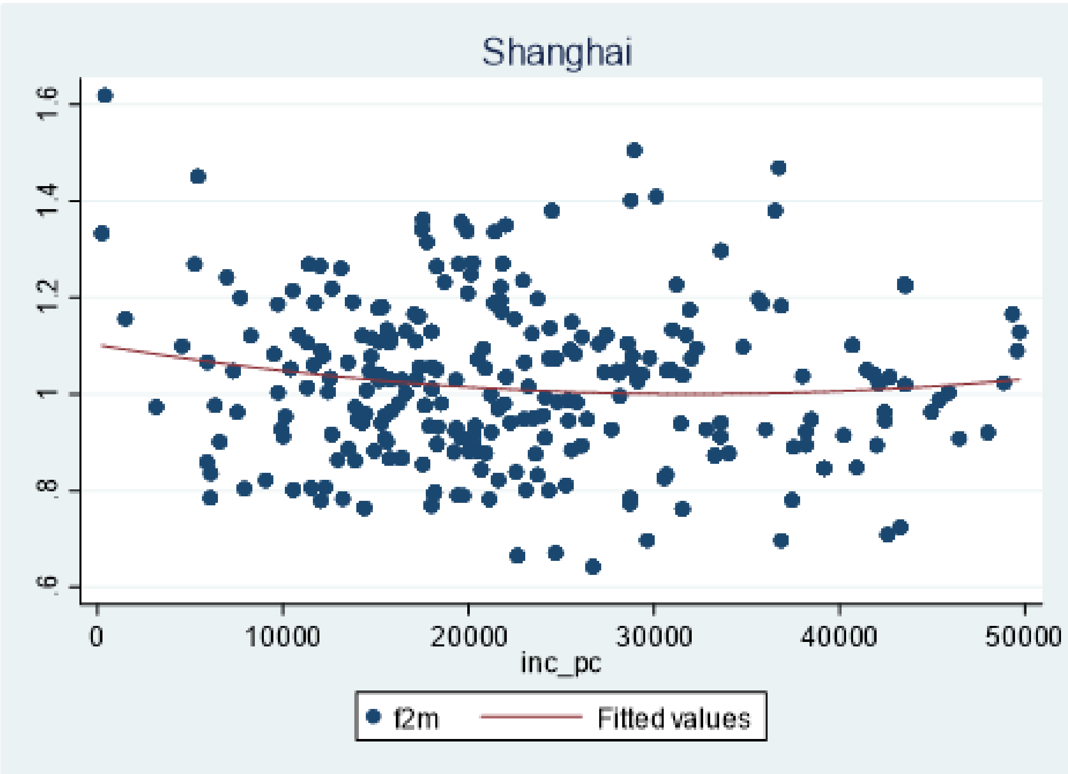

Liaoning
1.6
1.4
1.2
1
.8
.6
0
10000
20000
30000
40000
50000
inc_pc
f2m
Fitted values

(b)



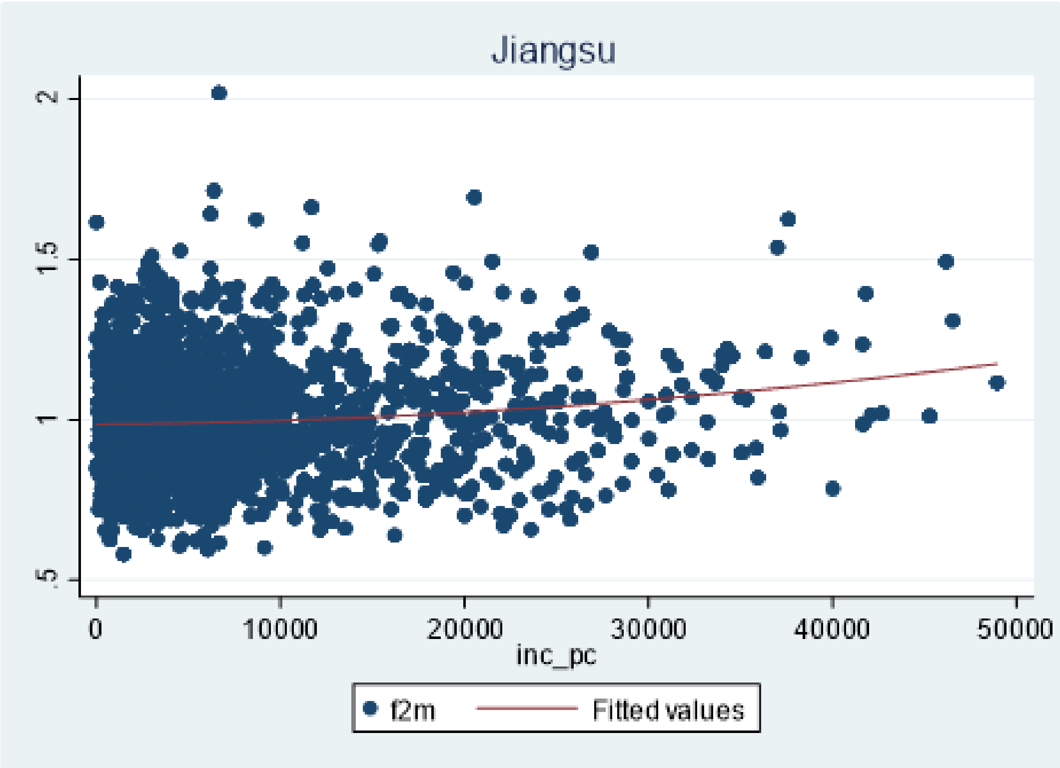
Heilongjiang
.5
1
1.5
2
0
10000
20000
30000
40000
50000
inc_pc
f2m
Fitted values

(c)

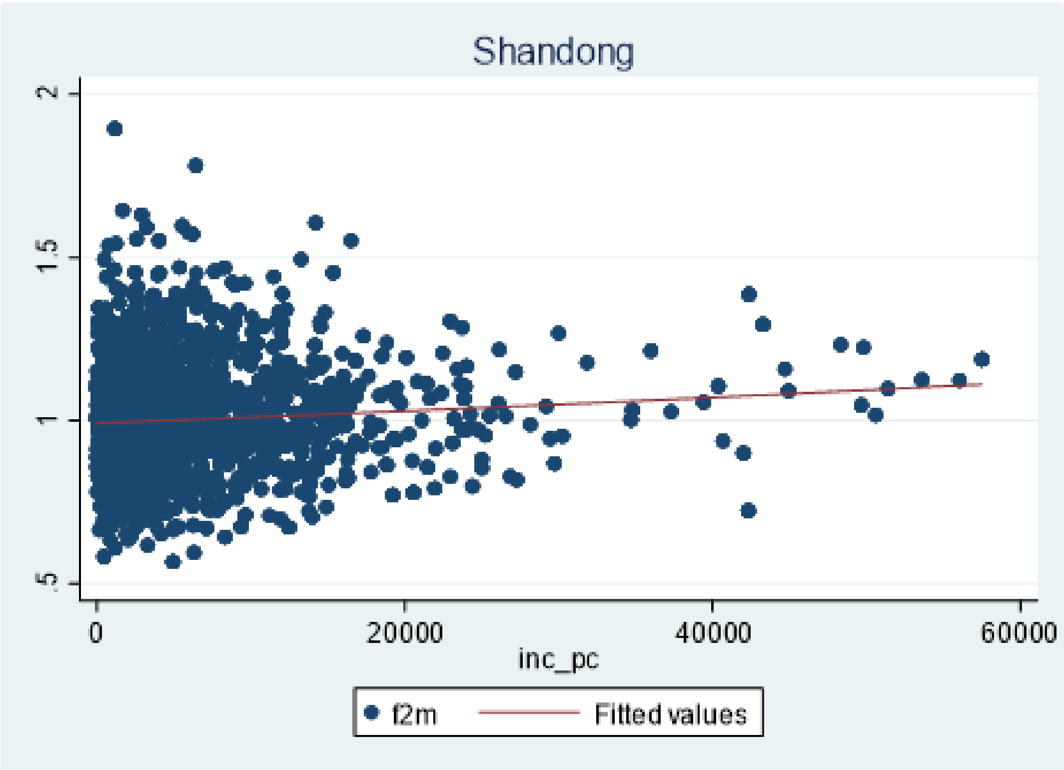


Shanghai
.6
.8
1
1.2
1.4
1.6
0
10000
20000
30000
40000
50000
inc_pc
f2m
Fitted values

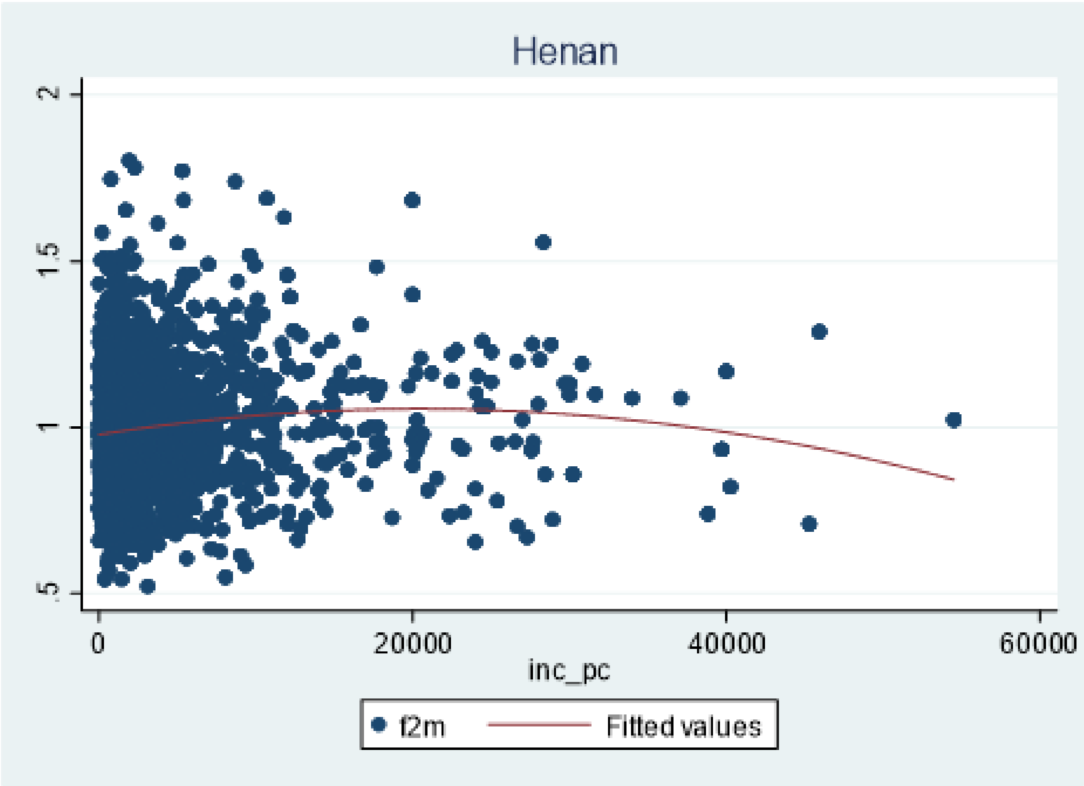
(d)


Jiangsu
2
1.5
1
.5
0
10000
20000
30000
40000
50000
inc_pc
f2m
Fitted values

(e)

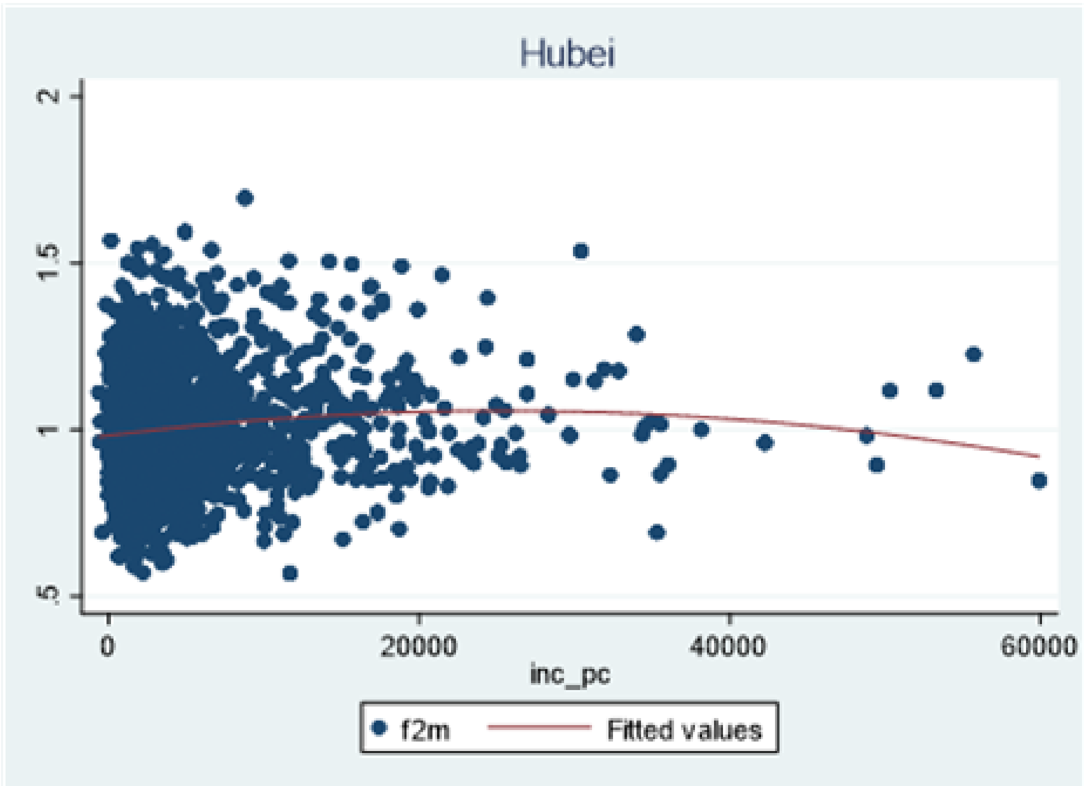

Shandong
2
1.5
1
.5
0
20000
40000
60000
inc_pc
f2m
Fitted values

(f)



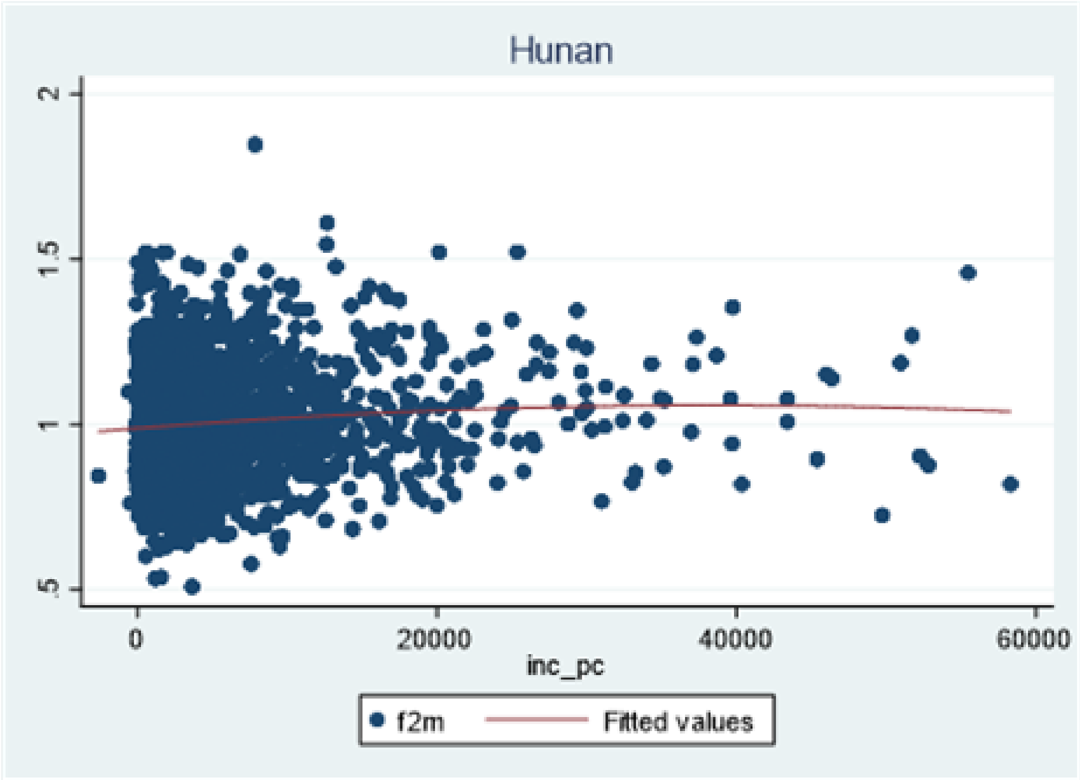
Henan
2
1.5
1
.5
0
20000
40000
60000
inc_pc
f2m
Fitted values

(g)

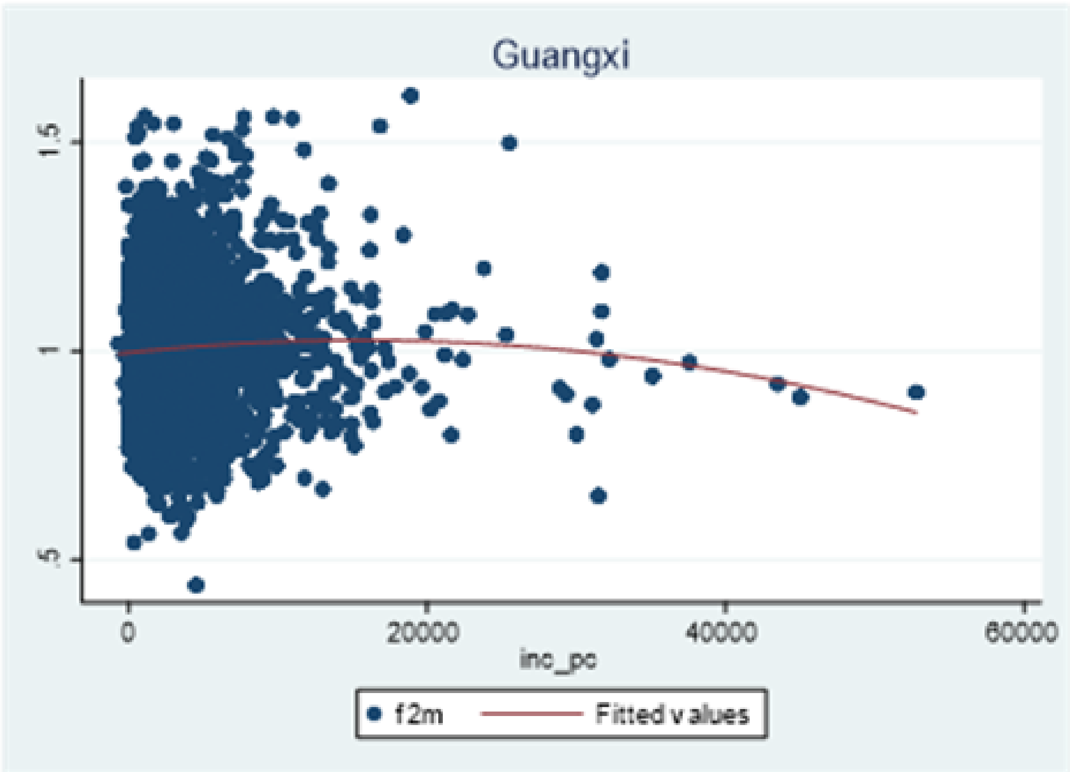


Hubei
2
1.5
1
.5
0
20000
40000
60000
inc_pc
f2m
Fitted values

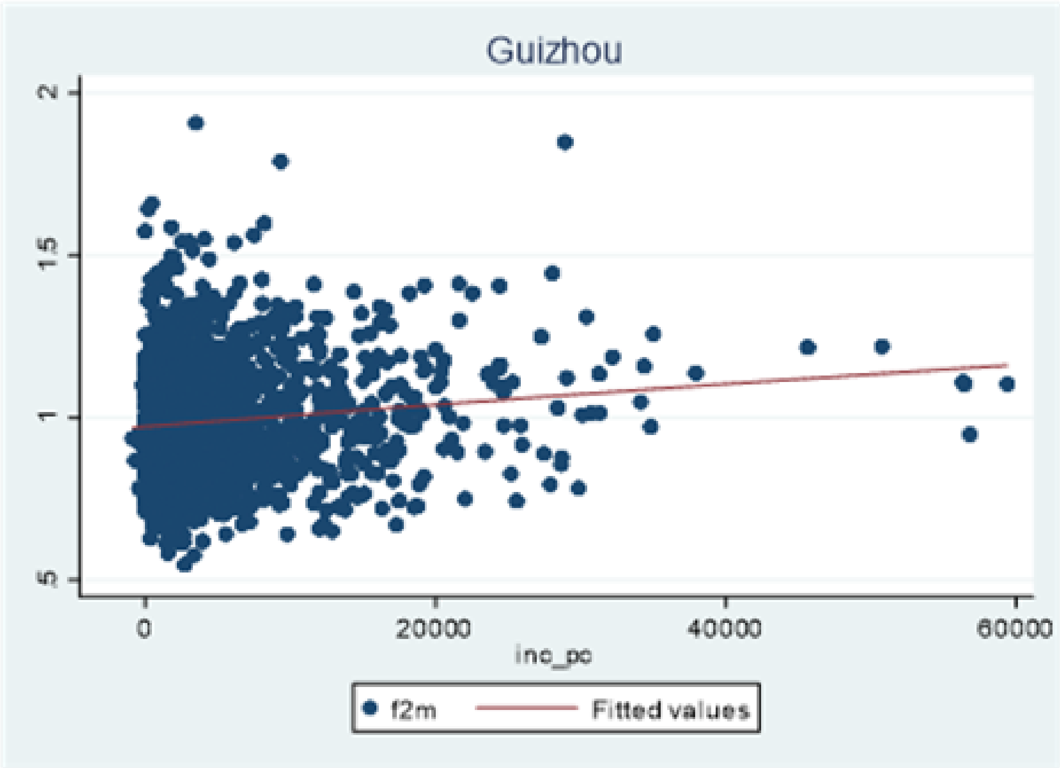
(h)


Hunan
2
1.5
1
.5
0
20000
40000
60000
inc_pc
f2m
Fitted values

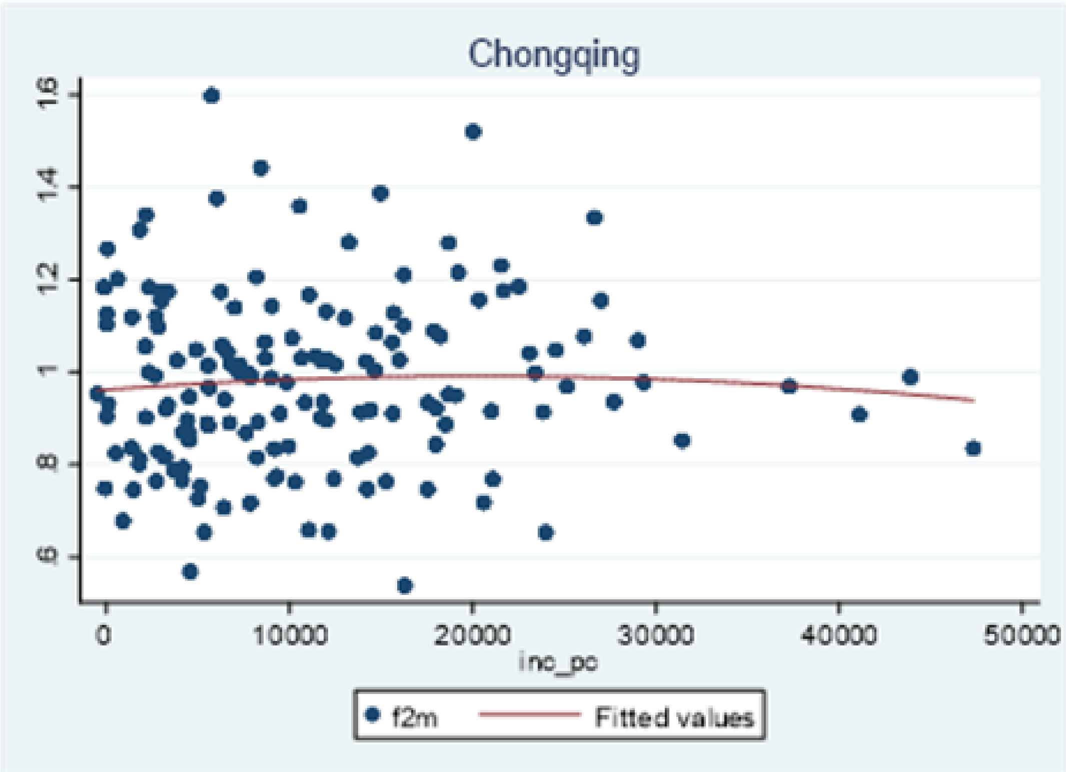
(i)


Guangxi
1.5
1
.5
0
20000
40000
60000
inc_pc
f2m
Fitted values

(j)



(k)



(l)

Figure III.9 (a)~(l). Male head to spouse BMI ratio ordered by per capita income for available years from 1898 to 2011.

Source: CHNS1989-2011

Similarly, father-to-mother BMI ratios do not show strong linear patterns as well-being indictors increase and they distribute evenly around a ratio of 1. Our finding is thus consistent with Sahn and Younger (2009).

Overall, deprivation among household members exists, but appears as a very weak relationship with well-being indicators. In example of father-to-child BMI ratios, half of the households in the sample show fathers have higher BMIs than children which suggesting fathers enjoy better health-related resources, while another half indicates the opposite.

d. Inequality Decomposition

I decompose overall inequality into intra-household inequality and between household inequality (Table III.5). Over half of the total health inequality is attributable to within household inequality. In 3 mega cities, Beijing, Shanghai, and Chongqing intra-household inequality rates are at around 60%. This result is consistent with previous literatures measuring the percentage of intra-household inequality. The result is empirically important because it demonstrates that omitting intra-household inequality will underestimate the real inequality by more than a half and hence largely impair the accuracy of inequality measurement and the effectiveness of policies.

Table III.5. Decomposition of total inequality into within versus between household inequality for available years from 1898 to 2011.

	between(%)	within(%)	between	within
Beijing	39.83%	60.17%	0.00412	0.00622
Liaoning	50.80%	49.20%	0.00551	0.00534
Heilongjiang	46.97%	53.03%	0.00462	0.00521
Shanghai	39.29%	60.71%	0.00385	0.00596
Jiangsu	47.35%	52.65%	0.00469	0.00522
Shandong	46.81%	53.19%	0.00465	0.00528
Henan	46.72%	53.28%	0.00485	0.00553
Hubei	51.07%	48.93%	0.00508	0.00487
Hunan	44.42%	55.58%	0.00413	0.00517
Guangxi	45.40%	54.60%	0.00408	0.00491
Guizhou	45.06%	54.94%	0.00413	0.00503
Chongqing	40.98%	59.02%	0.00457	0.00658
All provinces	50.21%	49.79%	0.00525	0.00520

Source: CHNS1989-2011

Conclusion and Discussion

This article explores intra-household health inequality by BMI. I find no evidence of intra-household or cross provincial Kuznets curve. I find no clear pattern of family member deprivation measured by BMI ratio, and no specific group (father, mother or children) is more advantaged in BMI measurement. Nevertheless, the intra-household inequality accounts for more than 50% of the overall BMI inequality.

BMI is an adequate measurement because it reflects individual level consumption of health-related food and non-food resources relative to personal need, and it also shares the advantages like non-zero positive value, easy collection and simple computation. However, using BMI has its limitations. Although I justified the reason of using BMI as well-being indicator in China (and other similar developing countries), it may be an inappropriate measurement in countries with high rate of obesity and food waste. One possible problem of using BMI as health indicator in China is the prevalence of women underweight. Studies show increasing fraction of Chinese women prefer slim body shape due to cultural norm and social environment, and adolescents in both genders appear to show relatively high percentage of underweight. (Chen and Shi, 2013). To the extent that more women choose to be slim, inequality in BMI does not necessarily reflect the inequality in access to the resources inside the family.

I do not find convincing evidence of a BMI Kuznets curve. The analysis applying CHNS data shows whether there exists a relationship (either quadratic or linear) between well-being measured in mean household BMI and intra-household inequality is uncertain. Our analysis shows no statistically significant correlation between individual BMI and income/expenditure per capita, which differs from Sahn and Younger (2009)'s finding which BMI is positively related to mean household BMI. Cross provincial relationship between inequality and well-being is inconclusive because of inadequate sample number. BMI ratios amongst household members show very weak relationship with well-being indicators, and they distribute almost symmetrically around ratio 1 as well-being conditions improve. The decomposition result proves intra-household inequality accounts for more than a half of total health inequality, thus suggests assuming equal household member well-being significantly underestimates the total inequality. Underestimation of real inequality should draw attention of policymakers because it could imply overestimating the effectiveness of inequality reduction policies or programs.

Policymakers should also not ignore intra-household health inequality because it could affect the possible outcomes of public transfer programs and increase of minimum wage.

The pattern of intra-household resource (re)allocation influence how much of the transfer or rise of minimum wage actually goes to the target individual. Sahn and Gerstle (2004)'s study show transfer payments to women affects family expenditure patterns. An interesting finding by Roemling and Qaim (2013) claim families with female household heads have significantly lower levels of nutritional inequality.

Further studies could develop on explaining the descriptive findings from this study, and discussing possible policies aiming to mitigate intra-household health inequality. The targets of such programs could be the least advantageous induvial without sufficient nutrition reflected in BMI inside households.

APPENDIX

```
. predict resid, residuals

. sktest resid

                    Skewness/Kurtosis tests for Normality
                                                          ------- joint -------
    Variable |        Obs  Pr(Skewness)  Pr(Kurtosis) adj chi2(2)   Prob>chi2
-------------+-----------------------------------------------------------------
       resid |    100,605     0.0000         0.0000           .            .
```

Figure III.A1. Normality test for OLS regression of intra-household BMI inequality on mean household BMI and mean household BMI square, all provinces pooled, for years available, 1989-2011.

```
. linktest

      Source |       SS           df       MS      Number of obs   =   100,605
-------------+----------------------------------   F(2, 100602)    =   2215.22
       Model |  .145170512         2  .072585256   Prob > F        =    0.0000
    Residual |  3.29639349   100,602  .000032767   R-squared       =    0.0422
-------------+----------------------------------   Adj R-squared   =    0.0422
       Total |    3.441564   100,604  .000034209   Root MSE        =    .00572

------------------------------------------------------------------------------
mean_log_d~n |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        _hat |  -.0340071   .0770014    -0.44   0.659    -.1849288    .1169146
      _hatsq |   114.8341   8.379871    13.70   0.000     98.40967    131.2586
       _cons |   .0021126   .0001746    12.10   0.000     .0017704    .0024547
------------------------------------------------------------------------------


Ramsey RESET test using powers of the fitted values of mean_log_deviation
       Ho:  model has no omitted variables
                 F(3, 100599) =     75.83
                      Prob > F =      0.0000

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of mean_log_deviation

         chi2(1)      =  4247.14
         Prob > chi2  =   0.0000
```

Figure III.A2. Model specification and homoscedasticity test for OLS regression of intra-household BMI inequality on mean household BMI and mean household BMI square all provinces pooled, for years available, 1989-2011.

REFERENCES

Bargain O, Donni O, Kwenda P. 2014. Intrahousehold distribution and poverty: Evidence from Côte d'Ivoire. Journal of Development Economics, 03/2014, Volume 107

Browning M, Chiappori PA. 1998. Efficient intra-household allocations: a general characterization and empirical tests. Econometrica 66(6): 1241–1278.

Browning M, Chiappori P and Weiss Y. 2014. Economics of the Family. Cambridge University Press.

Chen W, Shi Z. 2013. Trend in gender disparities of BMI and height between 2004 and 2011 among adolescents aged 17-18 years in Changzhou China. Asia Pacific Journal of Clinical Nutrition, 2013, Volume 22, Issue 3

Chen, Z., Eastwood, D., Yen, S. 2007. A decade's story of childhood malnutrition inequality in China: Where you live does matter. China Economic Review 18, 139-154.

Coates J, Patenaude B, and more. 2018. Intra-household nutrient inequity in rural Ethiopia. Food Policy, Volume 81

Deaton A, Grosh M. 2000. Consumption. In Designing Household Survey Questionnaires for Developing Countries: Lessons from Fifteen Years of the Living Standards Measurement Study, Vol. 1, Grosh M, Glewwe P (eds). World Bank: Washington, DC; 91–133.

del Ninno C, Dorosh P, Smit L, Roy DK. 2001. The 1998 flood in Bangladesh: disaster impacts, household coping strategies and response. IFPRI Research Report 122, International Food Policy Research Institute,Washington, DC.

Dunbar G, Lewbel A, Pendakur K. 2013. Children's resources in collective households: identification, estimation and an application to child poverty in Malawi. Am. Econ. Rev., 103 (2013), pp. 438-471

Growth reference data for 5-19 years. World Health Organization.

Haddad L, Alderman H, Appleton S, Song L, Yohannes Y, 2003. Reducing child malnutrition: how far does income growth take us? The World Bank Economic Review 17(1): 107–132.

Haddad L, Kanbur R. 1990."How Serious is the Neglect of Intra-Household Inequality?" Economic Journal, Vol. 100, pp. 866-881, September 1990.

Haddad L, Kanbur R. 1992. "Intrahousehold Inequality and the Theory of Targeting." European Economic Review, Vol. 36, Issue 2-3, pp. 372-378.

Haddad L, Kanbur R. 1993. "The Value of Intrahousehold Survey Data for Age-Based Nutritional Targeting" , Annales d'Economie et de Statistique, No. 29, pp. 65-81.

Haddad L, Kanbur R, Bouis H. 1995. Intra-household inequality at different welfare levels: energy intake and energy expenditure data from the Philippines. Oxford Bulletin of Economics and Statistics 57(3): 389–409.

Hadi AS. 1994. A modification of a method for the detection of outliers in multivariate samples. Journal of the Royal Statistical Society, Series (B) 56: 393–396.

Kanbur R. 2016. Intra-Household Inequality and Overall Inequality. Cornell Dyson School Working Paper 16-11, Cornell University, Ithaca.

Kuznets S. 1955. Economic growth and income inequality. American Economic Review 45(1): 1–28.

Lise J, Seitz S. 2011. "Consumption Inequality and Intrahousehold Allocations." The Review of Economic Studies, Vol. 78, No. 1, pp. 328-355.

McElroy MB. 1990. The Empirical content of Nash-bargained household behavior. Journal of Human Resources 25(4): 559–583.
Paukert F. 1973. Income distribution at different levels of development: a survey of evidence. International Labour Review 108(2): 97–125.
Pradhan M. 2001. Welfare analysis with a proxy consumption measure: evidence from a repeated experiment in Indonesia. Cornell Food and Nutrition Policy Program Working Paper 126, Cornell University, Ithaca.
Ravallion M. 2005. A poverty-inequality trade-off? Journal of Economic Inequality 3(2): 169–181.
Robinson RB, Cox CD, Odom K. 2005. Identifying outliers in correlated water quality data. Journal of Environmental Engineering 131: 651–657.
Roemling C, Qaim M. 2013. Dual burden households and intra-household nutritional inequality in Indonesia. Economics and Human Biology, Volume 11, Issue 4
Rothbarth, E. 1943. Note on a method of determining equivalent income for families of different composition. In Appendix IV in War-Time Pattern of Saving and Expenditure by Charles Madge, Occasional Paper No. 4. Cambridge: University Press.
Sahn D, Younger S. 2009. "Measuring Intra-Household Inequality: Explorations Using the Body Mass Index," Health Economics 18(S1): S13-S36.
Sahn D, Gerstle A. 2004. Child allowances and allocative decisions in Romania households. Applied Economics 36(14): 1513–1521.
Ssewanyana S, Younger SD. 2008. Infant mortality in Uganda: determinants, trends, and the millennium development goals. Journal of African Economies 17(1): 34–61.
The WHO Child Growth Standards. World Health Organization
Thomas D. 1990. Intra-household resource allocation: an inferential approach. Journal of Human Resources 25(4): 635–664.
Wagstaff, A., & Watanabe, N. (2003). What difference does the choice of SES make in health inequality measurement? Health Economics, 12(10), 885−890.